

## HW3 (100 POINTS)

### *Before you start:*

1. Download the zip file HW3.zip. This file will usually be saved in the Downloads folder (directory).
2. Move HW3.zip to your home directory by using GUI-drag and drop.
3. Open the zip file. A folder named HW3 will be created in your home directory.
4. Open a terminal session.
5. Change into your HW3 directory.
6. Use vi to make and edit files in directory HW3.
7. Open a second terminal by pressing **CMD+n** and use this terminal to test the Unix code that you write as answers to the questions.

**Follow these instructions in each file you submit**

- Include the class header
- Format of Q&A pairs

**Example:**

```
#mprocop2:01/30/2024:filename
```

```
#Q1
```

```
cat file
```

```
#Q2
```

```
ls -lt file
```

This HW has 3 parts, HW3-part1, HW3-part2 and HW3-part3, and will have 3 separate submissions on Gradescope.

All these parts are due Sunday Feb 11 11:59pm.

If you submit:

- HW3-part1 by Wed Feb 7 (11:59pm) will get 3 extra points.
- HW3-part2 by Friday Feb 9 (11:59pm) you will get 3 extra points.
- HW3-part3 by Saturday Feb 10 (11:59PM) you will get 2 extra points.

### **General Grading:**

All points for correct code

All points off for syntax errors

0.5 points for class header

0.5 points for correct Q&A format

-0.5 for case sensitive error

**Linux/Ubuntu and WSL users** please include in a comment line that you use Cygwin or Linux/Ubuntu in each script you submit. Write it down below the class header, like this:  
#class header  
#Ubuntu user

## HW3-part1 (40)

**1 (40)** Make a file called **Ex3-artists.bash** and answer the following questions which refer to the provided file *music-artists.dat*. **Answer each question with one line of code.**

The provided file *music-artists.dat* contains a list of classical music artists, and the fields are organized as follow:

- 1<sup>st</sup> field contains the death year
- 2<sup>nd</sup> field contains the last name
- 3<sup>rd</sup> field contains the first name
- 4<sup>th</sup> field contains the birth year

Look at the content of the file and see what the field separator is.

Q1. (8) Use the data file **music-artists.dat**.

Use command substitution to store in a variable called var, the artist's last name and first name, (2<sup>nd</sup> and 3<sup>rd</sup> fields) of the artist who died most recently. Your line of code should include a pipeline of commands, and command substitution. Notice that the output should be uppercase.

Variable var should store this output.

GRETCHANINOV:ALEXANDER TIKHONOVICH

Do not hardcode the output in your line of code. Suppose you do not know who died most recently. If you hardcode the output in your line of code, you will lose 4 points.

Moreover, if you literally store GRETCHANINOV:ALEXANDER TIKHONOVICH into the variable, you will lose all points.

Q2. (4) Use variable var in pipeline with a bash command to obtain to screen:

GRETCHANINOV - ALEXANDER TIKHONOVICH

Notice: you should substitute comma with space-space.

Q3. (8) Use the data file **music-artists.dat**.

In the file two artists' names contain typos. The last name BAC should be BACH and the last name RIESSSSS should be RIES

Your task is to **use both sed, tr and grep in pipeline** to fix the last name *BAC* by changing it to *BACH*, to squeeze repetitions of the letter *S* (in the name *RIESSSSS*), and finally to extract the two lines containing the correct information of these two artists which you will redirect into a file called `myartists.dat`. The file **myartists.dat** should contain this output:

```
1782:BACH:Johann Christian:1735
1838:RIES:Ferdinand: 1784
```

**You should generate this file with one line of code.**

Q4. (4) Use the data file **music-artists.dat**.

Extract the line containing the artist HELLER Stephen and append the output to the file **myartists.dat**. Notice that there are two artists named Stephen. Your task is to complete the code reported below by using a pipeline and redirection. **Do not use head and tail.**

```
grep Stephen music-artists.dat
```

Report the entire line of code in your answer, including the provided code.

Q5. (1) View the contents of the file **myartists.dat**. It should contain:

```
1782:BACH:Johann Christian:1735
1838:RIES:Ferdinand: 1784
1888:HELLER:Stephen:1813
```

Q6. (7) Use **input redirection with the tr command**, input the file **music-artists.dat**, and use a pipeline of commands to extract the first 6 columns of the first 5 lines. Notice the lowercase letters.

```
1782:b
1780:k
1915:g
1785:g
1836:w
```

Q7. (7) Run the script **script.bash**, which is within the HW3 directory (which should be your current directory), and redirect **ONLY** the standard error into the null device, and process the standard output to obtain to screen:

```
Today is
a beautiful day
```

Use one line of code with a pipeline. Do not use absolute path to run the script.

**Upload the following file to Gradescope [HW3-part1:](#)**  
**[Ex3-artists.bash](#)**

## HW3-part2 (34)

**1 (34)** The provided file **dow\_jones\_2011.dat** contains weekly stock data for Dow Jones Index during 2011, and comprises data reported by the major stock exchanges. View the contents of the file. The file contains 7 fields. The fields report the following:

- 1<sup>st</sup> field: yearly quarter (1 for Jan-Mar; 2 for Apr-Jun)
- 2<sup>nd</sup> field: the stock symbol (below a list of some stock symbols)

company	stock symbol
Bank of America	BAC
Intel	INTC
IBM	IBM
Johnson & Johnson	JNJ
Pfizer	PFE

- 3<sup>rd</sup> field: date, i.e., the last business day of the work week (usually Friday)
- 4<sup>th</sup> field: the price of the stock at the beginning of the week
- 5<sup>th</sup> field: the highest price of the stock during the week
- 6<sup>th</sup> field: the lowest price of the stock during the week
- 7<sup>th</sup> field: the price of the stock at the end of the week

Make a file called **Ex3-stock.bash** and answer the following questions. Follow Q&A format.

Q1.(8) Use awk to print the 1<sup>th</sup>, 2<sup>nd</sup>, 4<sup>th</sup> and 7<sup>th</sup> field of all the lines of the file **dow\_jones\_2011.dat** that contain the keyword 2 in the 1<sup>st</sup> field **AND** keyword JNJ **OR** keyword PFE in the 2nd field. **DO NOT USE THE CUT COMMAND IN BASH**

Q2.(8) Use awk on **dow\_jones\_2011.dat** and redirection to make a file called **diff\_week.dat**. This file should have: the 1<sup>st</sup> and 2<sup>nd</sup> field identical to the 1<sup>st</sup> and 2<sup>nd</sup> field of **dow\_jones\_2011.dat**, and the 3<sup>rd</sup> field as the weekly percentage price change of the stocks (i.e., percentage change at the end of the week with respect to the beginning of the week). Use the following formula:  $((7^{\text{th}} \text{ field} - 4^{\text{th}} \text{ field}) / 4^{\text{th}} \text{ field}) * 100$ .

Q3.(8) Use awk on **diff\_week.dat** to select the weekly percentage price changes (3<sup>rd</sup> field) that are greater than 7.0 and smaller than or equal to 10.0 and print only the 2<sup>nd</sup> and 3<sup>rd</sup> fields. Use printf in awk to format the output to obtain:

```
BAC  7.62
INTC 9.88
INTC 8.33
```

Q4.(1) Use sed on file **dow\_jones\_2011.dat** to change the field separator from a space to a colon and save the output in a file called **dow\_jones\_2011\_new.dat**.

Q5. (8) Use the file **dow\_jones\_2011\_new.dat** and a pipeline of commands to print just the highest price of stocks during 2011 (i.e., highest value in the 5<sup>th</sup> field), the stock symbol reporting the highest price (2<sup>nd</sup> field), and the date (3<sup>rd</sup> field) on which the highest price occurred.

Remember that now the field separator is a colon. Format the output like this by using `printf` in `awk`.

**Do not hardcode the answer in your code, and do not use command substitution.**

```
highest price=$173.5 stock=IBM date 5/6/2011
```

If you did not make **dow\_jones\_2011\_new.dat**, use the provided **dow\_jones\_2011\_new-back.dat**

Upload the following file to Gradescope [HW3-part2:](#)  
**Ex3-stock.bash**

## HW3-part3 (26)

1. (26) Make a file called **Ex3-misc.bash** and answer the following questions.  
Multiple data files are provided.

Q1. (4) The provided file **species.dat** contains information on amphibians and reptiles of the West Indies. The fields are the following:

- 1<sup>st</sup> field is the island name
- 2<sup>nd</sup> field is the island size
- 3<sup>rd</sup> field is the number of species

Before starting this exercise look at the content of the file.

The data set contains comment lines starting with `#`, and extra characters `*` in different places of the data file.

Your first task is to clean the data set, and make a clean file called `species-clean.dat`, i.e.: extract all the lines that do not contain the `#` sign, remove all the `*` characters, and save the clean output into a file called `species-clean.dat`. You need to achieve this with one line of code, by using bash commands in pipeline and a redirection operator.

If you do this in multiple lines of code, you will lose 1 point.

Q2. (5) Come up with some text manipulation to extract information from **species-clean.dat** by using at least 2 pipes, which means 3 commands. Write with a comment line the goal of your text manipulation. If you did not make the file species-clean.dat, you can use the provided file species-clean-bak.dat.

Q3. (8) Use the file **species-clean.dat**. If you did not make the file species-clean.dat, you can use the provided file species-clean-bak.dat.

Make this variable:

**var1=40**

Use the bash variable var1 in awk to print to screen the island names (1<sup>st</sup> field) in which the number of species (3<sup>rd</sup> field) is greater than the number stored in var1.

Hispaniola  
Cuba

Do not hardcode the island names.

Q4. (8) Calculate and print to screen the density of species on each island, i.e. the number of species divided by the island size (3<sup>rd</sup>/2<sup>nd</sup> field), sorted from the smallest density to the largest, and the island name uppercase. You should obtain this formatted output.

7.3e-04	CUBA
1.1e-03	HISPANIOLA
3.3e-03	JAMAICA
4.5e-03	PUERTO
8.7e-02	MONTSEERRAT
3.8e-01	SABA
1.0e+00	REDONDA

Notice the density is in scientific format, and there is more than one space between the two fields. You should add additional spaces by using the width. The number of spaces can be any, but it must be greater than 1.

Upload the following file to Gradescope **HW3-part3:**  
**Ex3-misc.bash**