# CS253
# Assignment 2 - Shell Scripting

IIT Kanpur

## Problem Statement

Through this assignment, you will acquire the knowledge and skills necessary to automate complex processes utilizing shell scripting. To complete this assignment, we encourage to use the **'sed'**, **'grep'** and **'awk'** commands to perform the tasks described below. Furthermore, you will be required to implement output redirection (>, >>, or equivalent methods), to write data to a file. You may consult the usage of the commands using the **'man'** command.

*Comments and code readability hold substantial weight in the evaluation..*

1. Write a script named **larger.sh** that takes two files as input. If either file does not exist, print a usage message telling the user how to use the script and exit. The script should read both input files line by line, count common prefixes for each line, and write the count to an output file. (Use loops, if conditions and functions to implement the logic, Do not use built-in commands that automate this process)

```
file1.txt:      file2.txt:
apple           apricot
banana          banter
cherry

2  # Common prefix of 'apple' and 'apricot' is 'ap'
3  # Common prefix of 'banana' and 'banter' is 'ban'
0  # Common prefix of 'cherry' and '' is none
```

2. Write a script named **main.sh** which takes 3 arguments.

   - The first file is an input file (`logs.csv`).
   - The second file is the output file where the results will be saved (Ex. `output.txt`).
   - The third file is a logs timestamp file in which the script records its execution steps (Ex. `log_ts.txt`).

The input file is a plain text file named **logs.csv** with the following structure

```
IP,timestamp,HTTP_method,request_path,status_code,content_length
192.168.1.1,2024-03-01 10:00:00,GET,/home,200,1024
192.168.1.2,2024-03-01 10:00:01,POST,/login,401,512
So on...
```

Use this file to perform the following tasks:

(a) Extract and print the unique IP addresses from the input file to the output file.

(b) Identify the top 3 HTTP methods (e.g., GET, POST, PUT, DELETE) used in the dataset. Save their names and frequencies to the output file.

(c) Find the total number of requests made in each hour of the day (0-23). Save the results to the output file in the following format:

```
Hour 0: 10 requests
Hour 1: 5 requests
...
Hour 23: 2 requests
```

3. Append essential steps timestamps of the script **main.sh** execution,s to the logs timestamp file (Ex. `log_ts.txt`). Use appropriate labels to differentiate between steps (e.g., "Input file exists", "Input file not detected", "Unique IP extraction completed", etc.).

Shell Command Example: `$ bash main.sh logs.csv output.txt log_ts.txt`

```
An example of the logs timestamp, saved in logs_ts.txt,
where the input file "logs.csv" is present:

> 2025-03-09 13:43:57 - Input file exists
> 2025-03-09 13:43:57 - Unique IP extraction completed
> 2025-03-09 13:43:57 - Top 3 HTTP methods identified
> 2025-03-09 13:43:58 - Hourly request count completed
> 2025-03-09 13:43:58 - Script execution completed
```

## Submission Instructions

- Create two shell script file named **larger.sh** and **main.sh** respectively.
- Include comments in your script to explain each section.
- You can use one or more awk script files or other auxiliary files if needed. Submit all source files in a zipped format.
- Assume the 'logs.csv' file is present in the same directory as the script.
- No need to submit any generated output/log files.

## Contact

Any queries regarding this assignment should be directed to **Aakashdeep Singh** (aakashdeepsingh@cse.iitk.ac.in). Please mention CS253 in the subject line while sending any email.