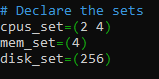
**Data Collection Script**

1. **Introduction**

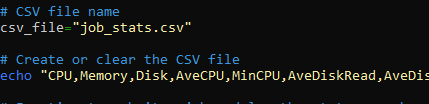
In this project, we are interested in learning how resource allocation affects the duration of the workflow along with the amount of resource usage. In the process of collecting system metrics data, we will run workflows by submitting them and monitoring the duration and usage of resources using the SLURM workload manager.

We will submit several workflows with several ways of resource allocation. After the workflows are complete, using SLURM accounting we will retrieve system metrics values and store them in a file which will be ready to be used as analysis material. The same process will be applied to all workflows. Therefore, we can save time doing the process by automating it using a shell script.

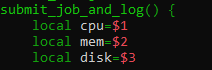
1. **Script**
2. Declare the sets that contain all possible resource requests.



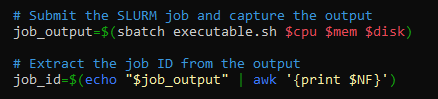
1. Create a csv file for logging. We add a header containing the system metrics fields (Table 1). These fields were selected to study CPUs, Memory, and IO.

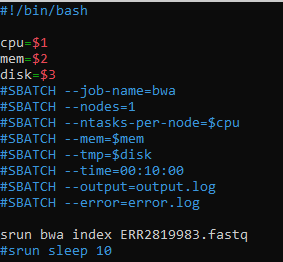


1. Define function to submit a job and log the system metrics

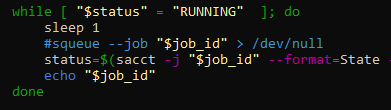
****

1. Submit the SLURM job and extract the job id

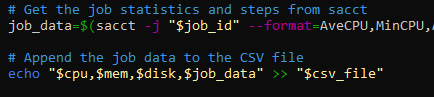




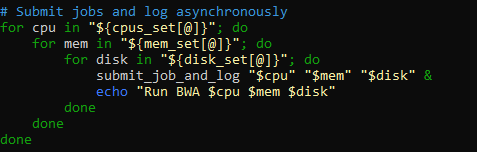
1. Wait for the job to finish by checking the job status every second.



1. Get the job system metrics by running sacct command and append the data to the CSV file.



1. Call the function with all possible resource requests



| Resource Usage | AveCPU |
| --- | --- |
| MinCPU |
| AveDiskRead |
| AveDiskWrite |
| AvePages |
| MaxDiskWrite |
| MaxPages |
| AveRSS |
| AveVMSize |
| MaxVMSize |
| Resource Allocation | AllocCPUS |
| AllocTres |

**Table 1. System metrics**