Lukasz Zajac lab 3 assignment

Task 1: Obtain the image set which will serve as input for processing

1. A set of 1000 images are available at the following path:
   /net/pr2/projects/plgrid/plgglscclass/image_data_sets/data/training_images/
2. Copy the images to a new directory in `$SCRATCH`.

```
[ares][plgqesterius@login01 plgqesterius]$ cp /net/pr2/projects/plgrid/plgglscclass/image_data_sets/data/training_
images/* -d images

[ares][plgqesterius@login01 plgqesterius]$
[ares][plgqesterius@login01 plgqesterius]$ ls
images  slurm_jobdir
[ares][plgqesterius@login01 plgqesterius]$ cd images
[ares][plgqesterius@login01 images]$ ls
vid_4_10000.jpg  vid_4_13820.jpg  vid_4_17840.jpg  vid_4_21420.jpg  vid_4_26320.jpg  vid_4_500.jpg
vid_4_1000.jpg   vid_4_13840.jpg  vid_4_17860.jpg  vid_4_21440.jpg  vid_4_26340.jpg  vid_4_5080.jpg
vid_4_10020.jpg  vid_4_13860.jpg  vid_4_17880.jpg  vid_4_21460.jpg  vid_4_26360.jpg  vid_4_5100.jpg
vid_4_10040.jpg  vid_4_13880.jpg  vid_4_17900.jpg  vid_4_21480.jpg  vid_4_26380.jpg  vid_4_5120.jpg
vid_4_10060.jpg  vid_4_13900.jpg  vid_4_17920.jpg  vid_4_21500.jpg  vid_4_26400.jpg  vid_4_5140.jpg
vid_4_10080.jpg  vid_4_13920.jpg  vid_4_17940.jpg  vid_4_21520.jpg  vid_4_26420.jpg  vid_4_5160.jpg
vid_4_10100.jpg  vid_4_13940.jpg  vid_4_17960.jpg  vid_4_21540.jpg  vid_4_26440.jpg  vid_4_5180.jpg
vid_4_10120.jpg  vid_4_13960.jpg  vid_4_17980.jpg  vid_4_21560.jpg  vid_4_26460.jpg  vid_4_5200.jpg
vid_4_10140.jpg  vid_4_13980.jpg  vid_4_18000.jpg  vid_4_21580.jpg  vid_4_26480.jpg  vid_4_520.jpg
vid_4_10160.jpg  vid_4_14040.jpg  vid_4_1800.jpg   vid_4_21600.jpg  vid_4_26500.jpg  vid_4_5220.jpg
```

Task 2: Create `converter.sh` script for image conversion using ImageMagick

1. The conversion command using ImageMagick is:

```
magick convert -adaptive-resize 3840x2160 -adaptive-sharpen 10 <input> <output>
```

2. Each image should take about 30 seconds to process.
3. The script should:
   Be done in Bash or any other scripting language.
   Take imagename as an argument. The output can be determined based on the input file or provided as a second parameter.
4. It is best to test the conversion inside of an interactive job to avoid stressing the login nodes.
   Refer to Lab02 for an example on how to start an interactive job in the plgrid-now partition.

converter.sh

```
magick convert -adaptive-resize 3840x2160 -adaptive-sharpen 10 $SCRATCH/images/$1 output/$1
```

Task 3: Create a job script and execute it, the job should:

1. Process in parallel multiple images inside of a single job
2. Use the converter script from previous point
3. Parallelization should be done using xargs ... srun combination
4. Determine the proper –P parameter for xargs
5. Determine the proper srun arguments
6. Each process should be a single task using 1 core
7. How to create a convenient input for xargs?
8. Please add the:

```
export OMP_NUM_THREADS=1
```

right after the #SBATCH directives, this will instruct imagemagic to use single core per process.

9. The job template is included below

```
#!/bin/bash -l
#SBATCH --nodes=1
#SBATCH --ntasks-per-node=48
#SBATCH --cpus-per-task=1
#SBATCH --account=plgglscclass24
#SBATCH --partition=plgrid-now
#SBATCH --time=01:00:00
# above config is mandatory!

export OMP_NUM_THREADS=1
# modules initialization
...
# data handling and work
...
# end of the script'
```

SOLUTION

```
#!/bin/bash -l
#SBATCH --nodes=1
#SBATCH --ntasks-per-node=48
#SBATCH --cpus-per-task=1
#SBATCH -A plglscclass24-cpu
#SBATCH --partition=plgrid-now
#SBATCH --time=01:00:00
# above config is mandatory!

export OMP_NUM_THREADS=1
module load imagemagick
cat files | xargs -P 48 -L 1 -n 1 srun --cpus-per-task=1 bash ./converter.sh
```