

New Interactive Job Behavior after SLURM Update

The behavior of interactive jobs has changed after last week's update to SLURM's latest release. Previously, when requesting a GPU in an interactive job, an additional `srun` command was required to use the GPU.

```
$ salloc -n 1 --gres=gpu:1
salloc: Granted job allocation 16806593
salloc: Waiting for resource configuration
salloc: Nodes csn-037 are ready for job
$ srun --gres=gpu:1 nvidia-smi
Mon Mar  8 12:58:08 2021
+-----+
| NVIDIA-SMI 450.51.05      Driver Version: 450.51.05      CUDA Version: 11.0      |
+-----+-----+-----+-----+-----+-----+
| GPU  Name           Persistence-M| Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp   Perf   Pwr:Usage/Cap|      Memory-Usage | GPU-Util  Compute M. |
|                                           MIG M.         |
+-----+-----+-----+-----+-----+-----+
|   0   Tesla K20m          On          | 00000000:83:00.0 Off |             0         |
| N/A   29C    P8         26W / 225W |  0MiB /  4743MiB |      0%      Default  |
|                                           N/A              |
+-----+-----+-----+-----+-----+

+-----+
| Processes:                                     |
|  GPU   GI    CI          PID    Type    Process name                      GPU Memory |
|      ID   ID                                     |            Usage                    |
+-----+-----+-----+-----+-----+
| No running processes found                    |
+-----+
```

This additional `srun` is no longer required. The allocated GPU can be used immediately.

```
$ salloc -n 1 --gres=gpu:1
salloc: Granted job allocation 16806705
salloc: Waiting for resource configuration
salloc: Nodes lac-342 are ready for job
$ nvidia-smi
Mon Mar  8 13:00:08 2021
+-----+
| NVIDIA-SMI 450.51.05      Driver Version: 450.51.05      CUDA Version: 11.0      |
+-----+-----+-----+-----+-----+-----+
| GPU  Name           Persistence-M| Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp   Perf   Pwr:Usage/Cap|      Memory-Usage | GPU-Util  Compute M. |
|                                           MIG M.         |
+-----+-----+-----+-----+-----+-----+
|   0   Tesla K80          On          | 00000000:0B:00.0 Off |             0         |
| N/A   42C    P8         29W / 149W |  0MiB / 11441MiB |      0%      Default  |
|                                           N/A              |
+-----+-----+-----+-----+-----+

+-----+
| Processes:                                     |
|  GPU   GI    CI          PID    Type    Process name                      GPU Memory |
|      ID   ID                                     |            Usage                    |
+-----+-----+-----+-----+-----+
| No running processes found                    |
+-----+
```

The original method will still work, so workflows that depends on running additional `srun` commands within an allocation—such as testing job steps to later be submitted in a batch job—will not need to be adjusted.

If you have any questions about this change, please contact us at <https://contact.icer.msu.edu>.