

Fine tune your workload

The web portal provides data for you to understand the actual utilization of your jobs, running or completed.

When you have a job running, click jobs, then the running jobs from the top menu bar to see the below screen. This screen tells the duration and consumption/allocation status of memory and CPU. If your job is underutilized, you may consider canceling it by clicking the cancel link on this screen and re-running it with a lower setting.

Job id	Account	Partition	Duration	Memory	Cpu
1218	appcara	batch	18s	0.4 / 7.8 GB (4.80%)	10s / 1m12s (13.89%)

After a job is completed, you may want to understand the utilization. Open the completed jobs window, search and click on the job id to bring up the below details window. In the efficiency section, you can find the allocated CPU time and memory, with the actual consumed figures.

Info			
Name:	hc.temp.sbatch	State:	failed
Account:	appcara	User:	loki
Partition:	gpu	Submitted at:	2022.11.12 12:09
Started at:	2022.11.12 12:09	Ended at:	2022.11.12 12:17
Duration:	8m7s	Suspended:	0s
Inductive charge			HKD \$0.589
Node	CPU Oneasia	GPU Oneasia	
oaklbhpcgpu001	0.30	0.04	
Total	0.30	0.04	
Unit rate	0.14	12.74	
Charge	0.04	0.55	
Efficiency			
Allocated CPU time:	16m14s	Allocated Memory:	4,000 MB
Total CPU time:	1m26s	Max RSS.:	257 MB
CPU efficiency:	8.9%	Memory efficiency:	6.4%
Submit line			
sbatch /pfss/home/Loki/hc.temp.sbatch			

There is also an inductive charge section. Telling what nodes have been allocated to this job and how much the charge is. Here is the standard cost, but not including the tiered price and discount.

To analyze the GPU utilization status of your job, you may want to profile your application. The cluster has both NVIDIA Visual Profiler and Nsight Compute installed. We provide them in both Lmod and through the NVHPC container. Please check them out if needed.

Revision #5

Created 4 November 2022 07:43:19 by Loki Ng

Updated 13 November 2022 16:51:54 by Loki Ng