ISBC

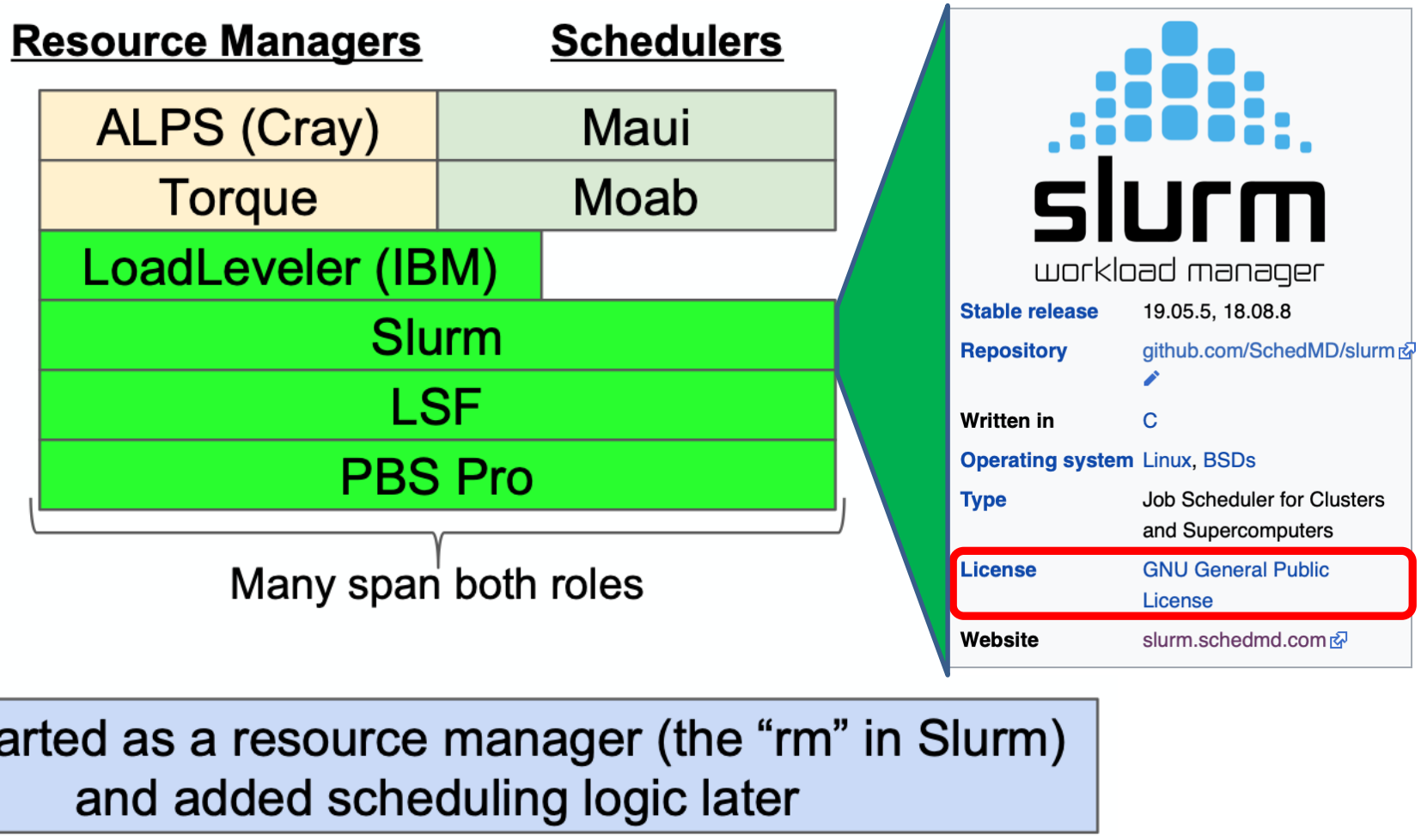
Slurm Resource Manager Service for GPGPU Computing

WanHee, Kim | kimwh@isbc.co.kr | ISBC Inc 2019

2020-02-13

Slurm is one of Good Resource Manager and under GPL

End-User
- KISTI
- LGE
- Intel RnD
- Brigham Young University
- Havard University
- LLNL Research
- Meteo France
- MIT
- NASA Center
- Swiss National Supercom
- NERSC
- Facebook
- Barcelona HPC Center
- Berkeley Lab
- Google
- AWS



Slurm started as a resource manager (the “rm” in Slurm) and added scheduling logic later

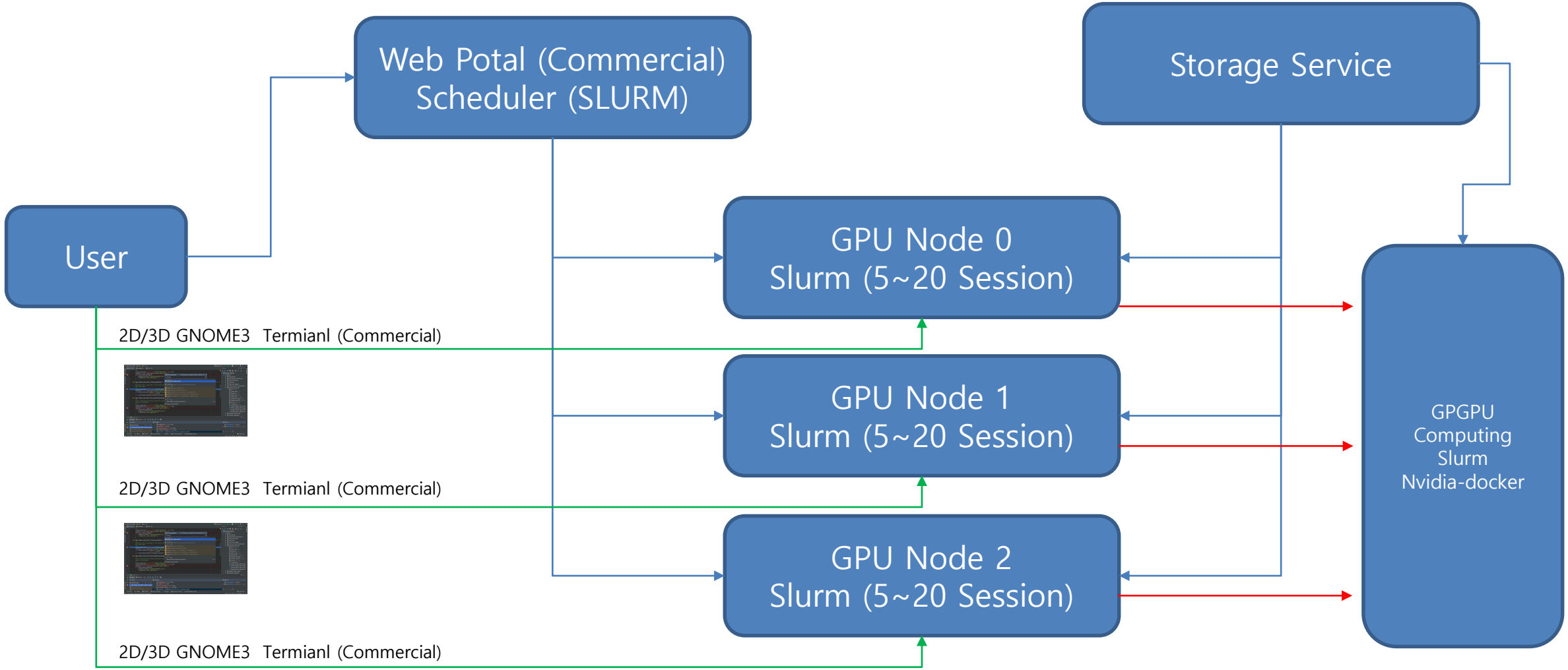
Comparison LSF vs SLURM

	Spectrum LSF	SLURM
Implementation Language	C/C++	C
Authentication	Multiple - OS Authentication/Kerberos	Munge, None, Kerberos
Heterogeneous exec node	Heterogeneous HW/OS (AIX, Linux, Windows)	Heterogeneous HW
Job Priority	Policy based – No queue to compute node binding	Yes
Group Protiry	Policy based – No queue to compute node binding	Yes
Queue Type	Batch , Interactive, Checkpointing, parallel and combination	Multifactor fairshare
SMP aware	Yes and GPU aware	Yes , Support GPU
Max exec node	> 9000 compute hosts	Tested 120K nodes
Max Job Submit	> 4 Mio job a day	Tested 100K Jobs
CPU scavenging	Yes	No
Parallel Job	Yes	Yes
Job Check Point	Yes	Yes
License	Proprietary (per core / per socker based charge?)	GPL v2

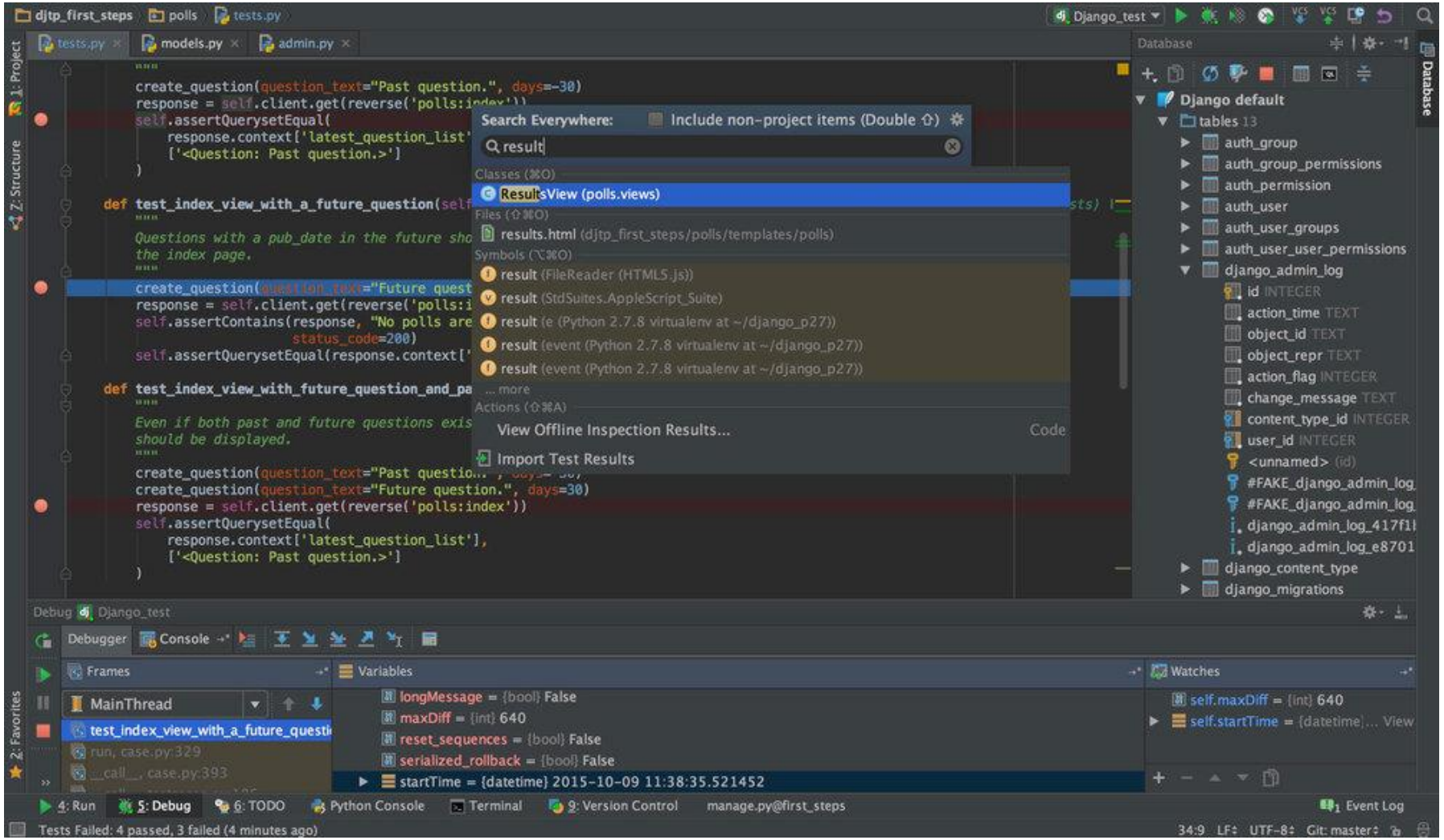
Key Feature of SLURM for GPGPU

- Scales to millions of cores and tens of thousands of GPGPUs
- Military grade security
- Heterogenous platform support allowing users to take advantage of GPGPUs.
- Flexible plugin framework enables Slurm to meet complex customization requi rements
- Topology aware job scheduling for maximum system utilization
- Open Source
- Extensive scheduling options including advanced reservations, suspend/resum e, backfill, fair-share and preemptive scheduling for critical jobs
- No single point of failure

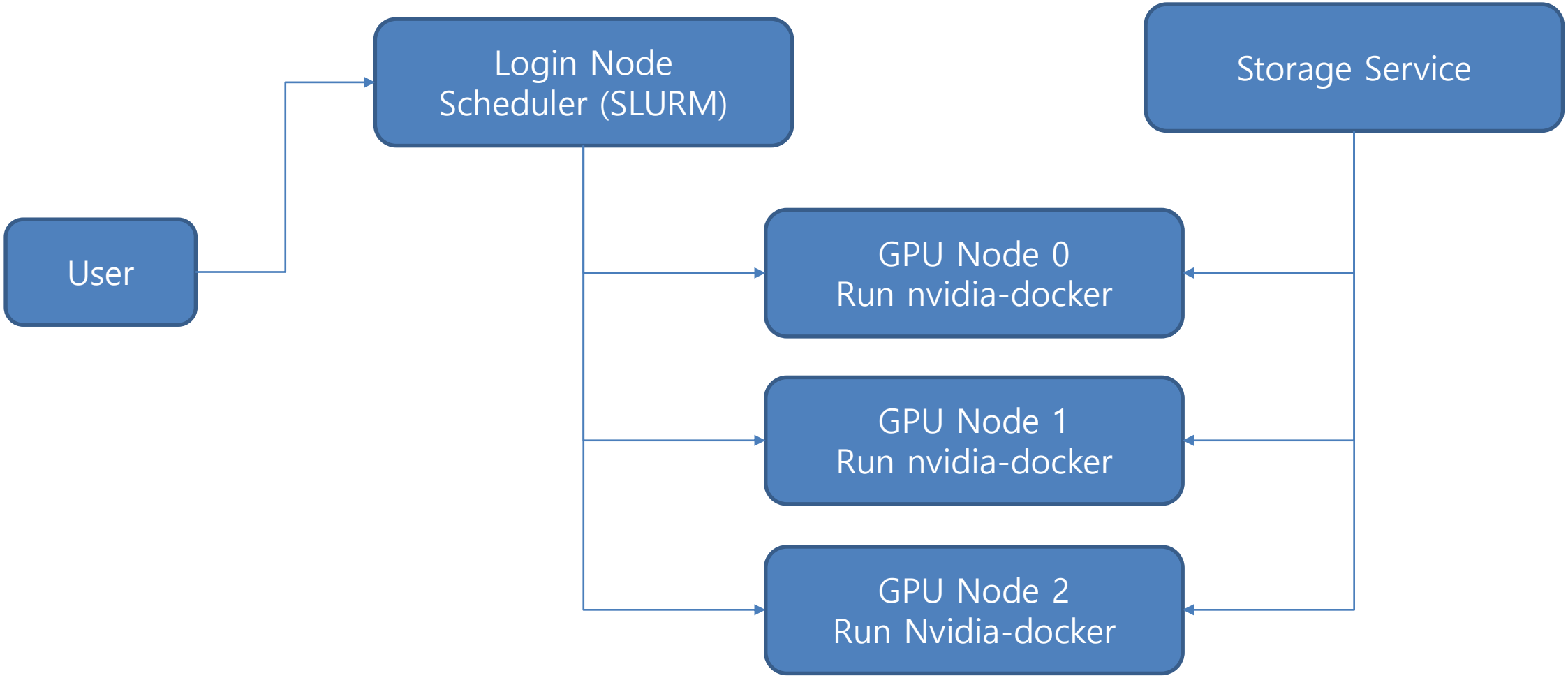
Uses Case : Interactive Job 2D/3D Session on GPU Nodes (For Development) – Pre/Post



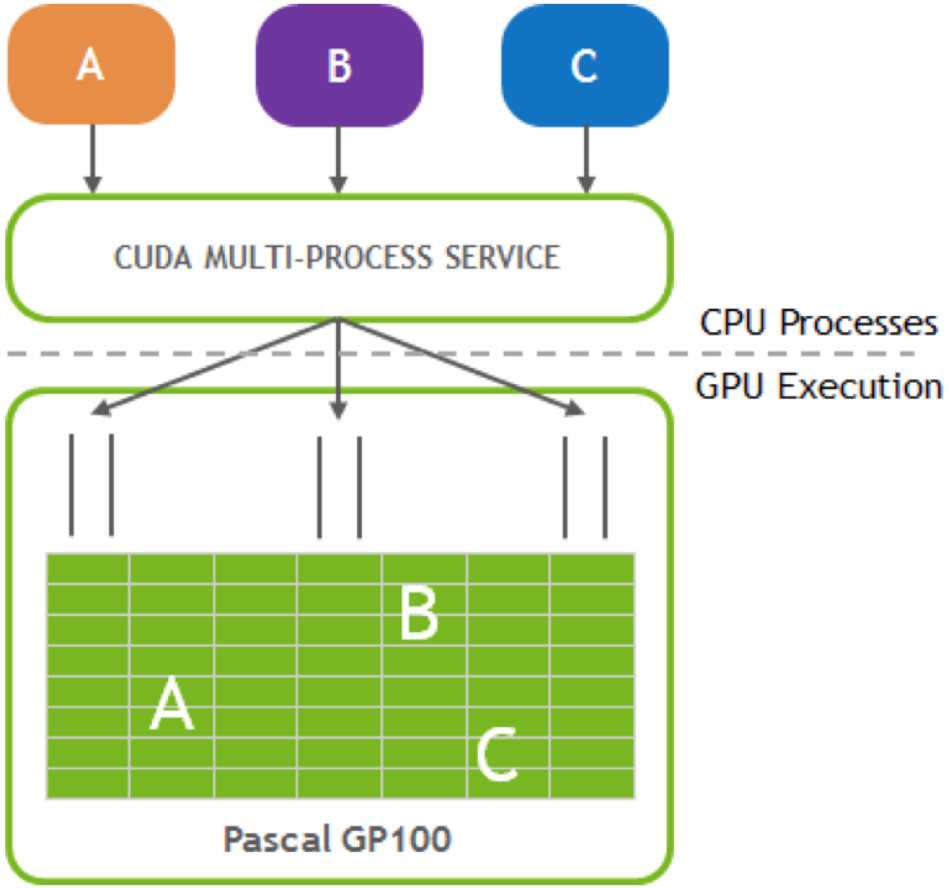
Uses Case : Interactive Job 2D/3D Session on GPU Nodes (For Development) pycharm

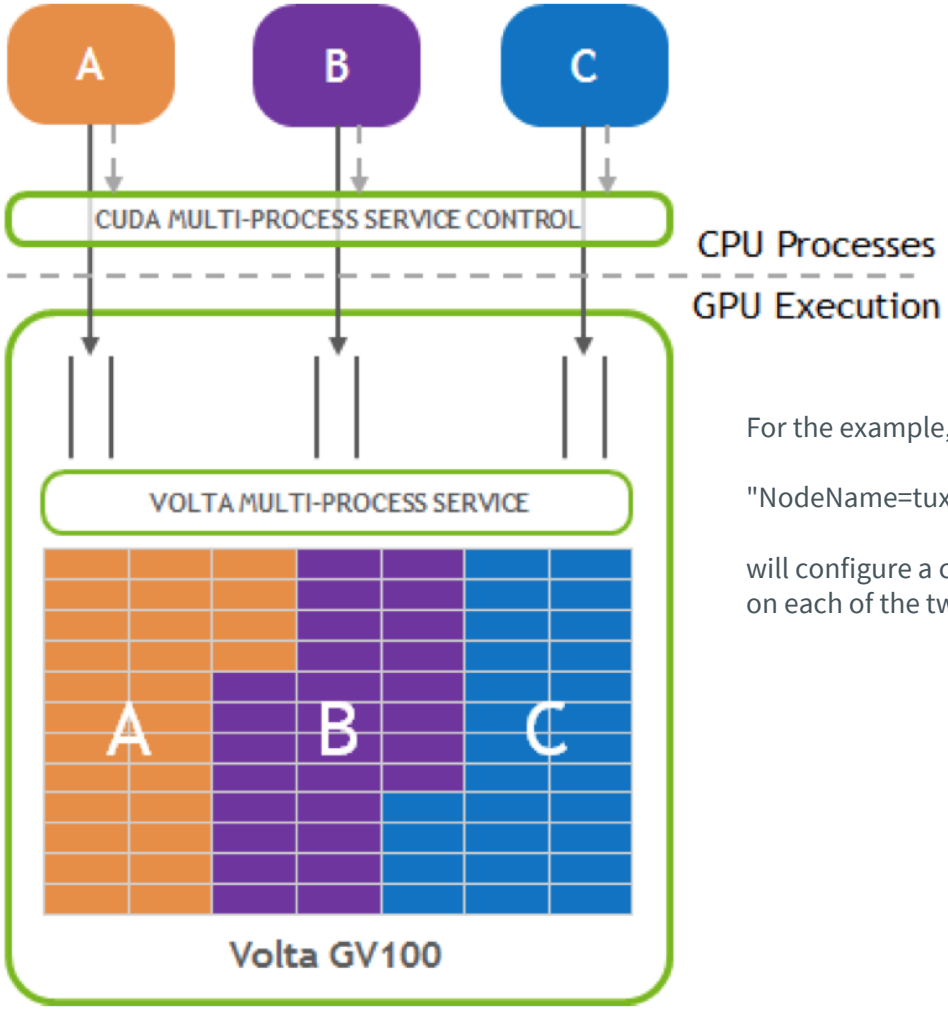


Uses Case : Login Nodes Run CUDA Application – GPGPU Computing



Slurm support nvidia mps control (P100 vs V100)



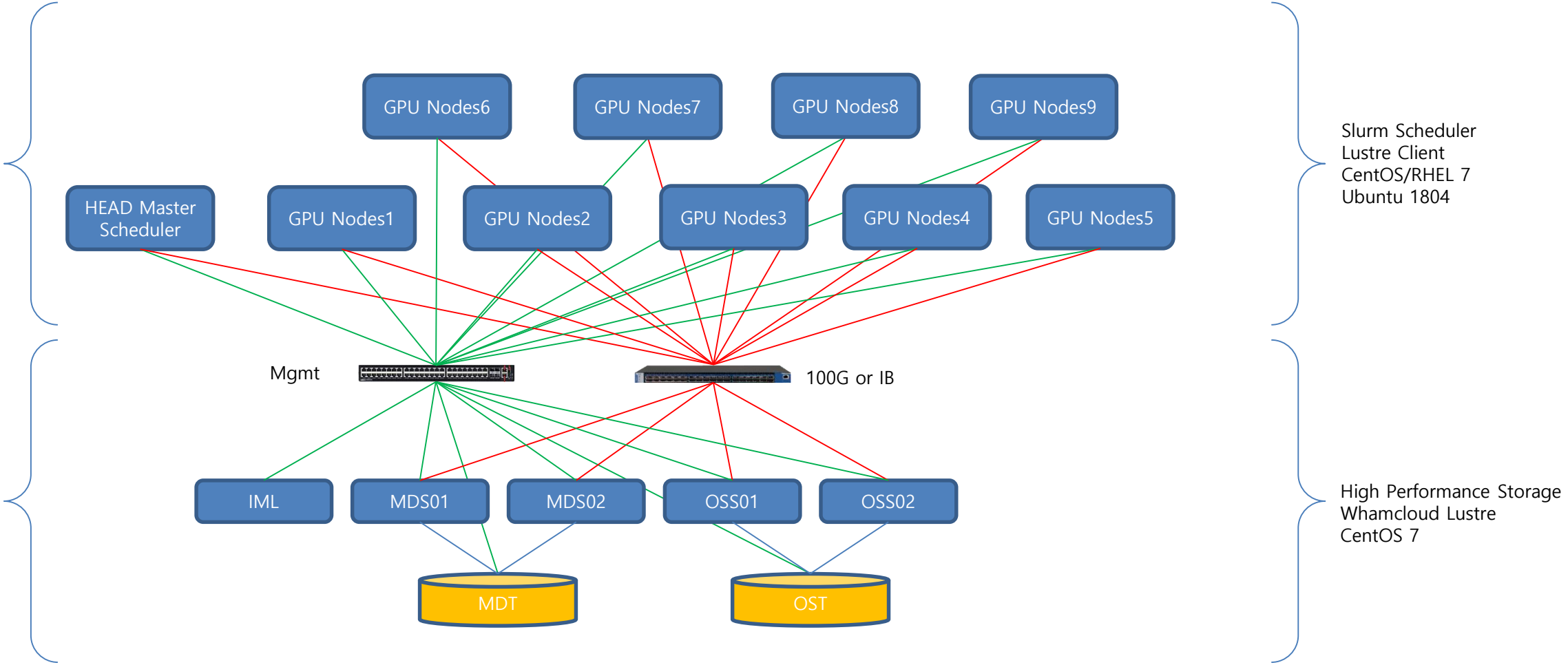


For the example,

"NodeName=tux[1-16] Gres=gpu:2,mps:200"

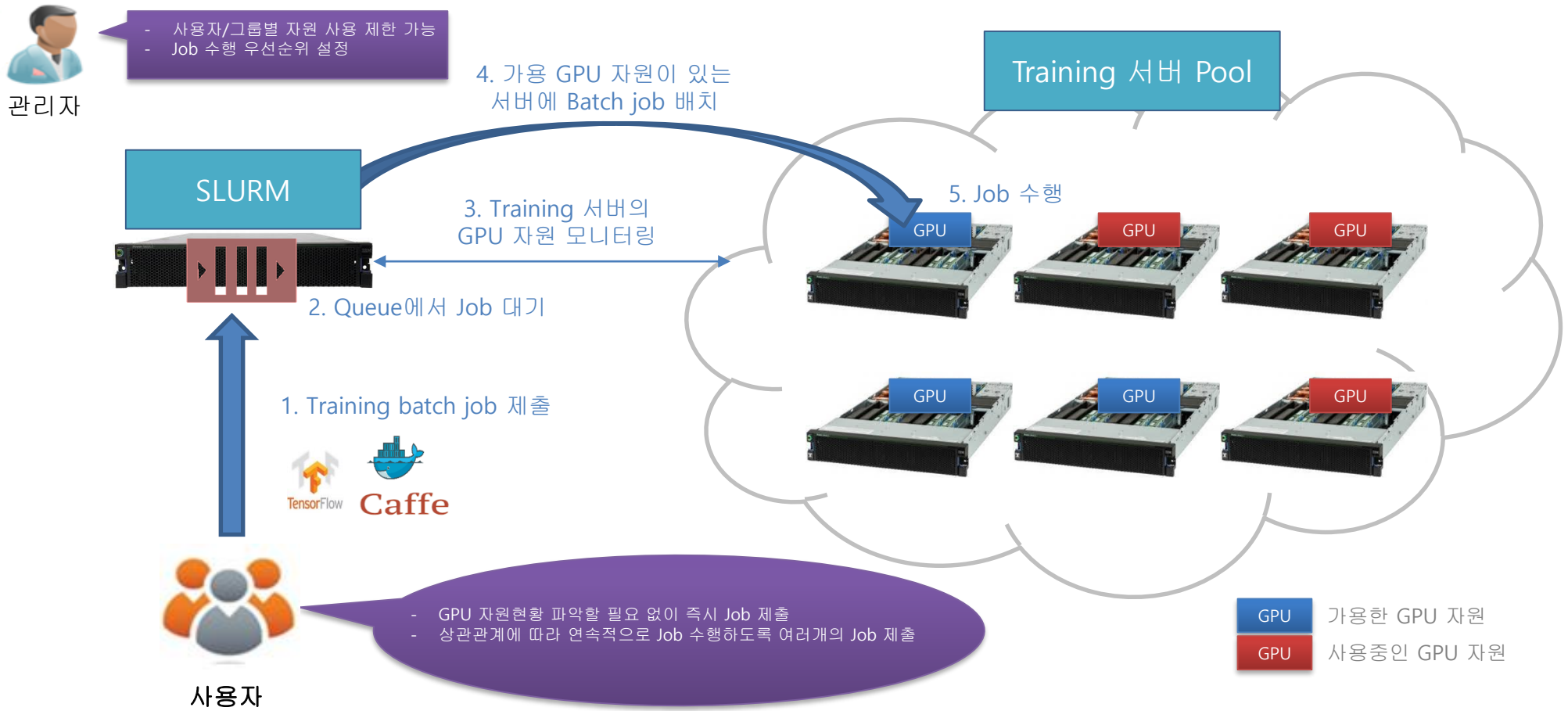
will configure a count of 100 gres/mps resources on each of the two GPUs.

ISBC GPU AI Cloud Service



GPU Resource Management and Scheduling

- 다수 사용자의 Training Job을 적절한 GPU 서버에 배치하는 SLURM
  - SLURM은 GPU Pool의 자원 사용을 모니터링하고, 다수의 사용자가 제출하는 Job을 정책에 따라 가용한 GPU 서버에 자동으로 배치, 수행시킵니다.



Appendix : GPGPU Monitoring Env screen shot


Q&A


THANK
YOU
MERCI
DANKE
GRAZIE
BEDANKEN
¡GRACIAS!