

Cloud Bursting with Secure Staging / GPU Burst Buffer with GPU/NVMe Direct

Over-Petascale
Universal Supercomputer

1. Cloud Bursting with Secure Staging

Problem

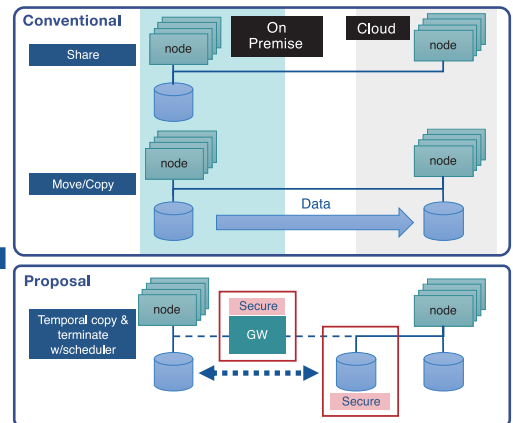
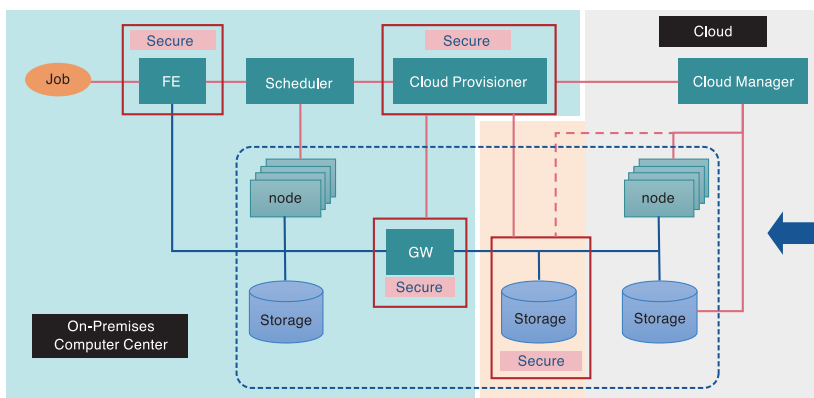
- Data must be shared between cloud and on-premises computer center.
- Sharing on-premises storage may degrade performance and security level.
- Some users have security concern about leaving data on shared storage in cloud.

Proposal : Staging data just in time for Cloud Bursting

Operation

1. Job input
2. Resource reservation (node, storage, network)
3. Secure stage in
4. Job execution
5. Secure stage out
6. Resource release

System Architecture



2. GPU Direct Burst Buffer

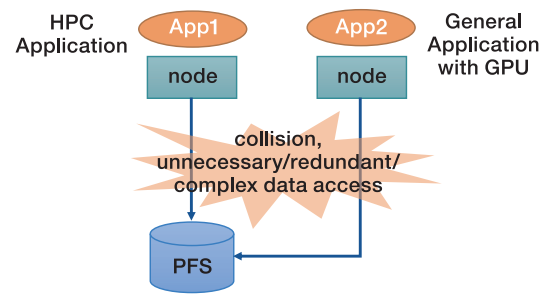
Problem

- Delay in cascading/interacting multiple application/process. e.g. visualization/application integration

Proposal : GPU Burst Buffer

- Cache with direct transfer between GPU and NVMe/NVMe.
- 2nd application executed on GPU Burst Buffer.

Conventional



Proposal

Option: If application 2 is heavily CPU bound

