Cybermedia Center

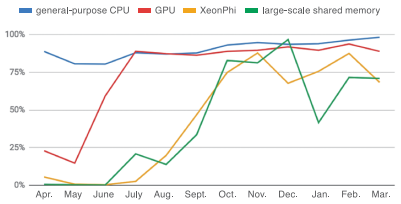
Osaka University, Japan

SC19 BOOTH 1843

Feasible Study of Cloud Bursting on OCTOPUS

Background

OCTOPUS is a hybrid cluster system of general-purpose CPU nodes (Skylake), many-core nodes (Knights Landing), GPU nodes (Tesla P100) and large-scale shared memory nodes with a 2PB Lustre storage (DDN EXAScaler). Since we started the operation of OCTOPUS, OCTOPUS has kept a higher utilization ratio. In particular, CPU and GPU nodes have a tendency of being demanded all year round. As a result, user waiting time is becoming longer.



Goal and Purpose

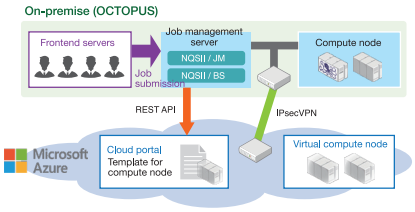
- Offloading the workload on OCTOPUS to the Cloud to alleviate the peak in hope that
  - User waiting time is reduced.
  - Higher throughput is realized.
  - We do not receive any complaint about waiting time (Higher satisfaction is achieved).
- Investigating the feasibility for the future integrated use of our supercomputing systems with the cloud.
  - For scaling out in need of compute resources.
  - For deploying and delivering the brand-new processors and accelerators to our user scientists and researchers.

OCTOPUS-Azure Environment

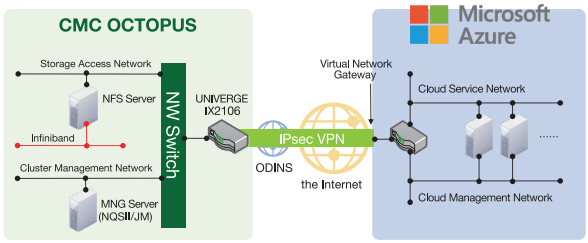
The right figure shows the overview of the first implementation of OCTOPUS-Azure environment where the workload on our on-premise environment is offloaded to the cloud when the demand for computing capacity spikes. For this study, we have introduced Microsoft Azure as an IaaS cloud to be integrated with OCTOPUS.

For realizing this environment, the following three have been considered.

1. Cloud-bridge network.
2. Virtual compute node deployment.
3. Job manager.



Cloud-bridge network



IPSec VPN has been established between on-premise and cloud.

- Lustre on OCTOPUS have to be accessed from Azure.
- NQSII/JM, the job server on OCTOPUS have to communicate with virtual compute node.

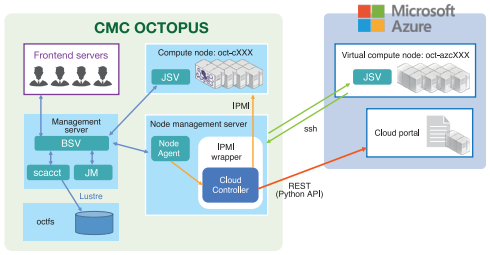
Virtual compute node deployment

Size	vCPU's	Memory:GiB
Standard_F2s_v2	2	4
Standard_F4s_v2	4	8
Standard_F8s_v2	8	16
Standard_F16s_v2	16	32
Standard_F32s_v2	32	64
Standard_F48s_v2	48	96
Standard_F64s_v2	64	128
Standard_F72s_v2	72	144

Theoretical Computing Speed	1.463 PFLOPS	
Compute Node	General purpose CPU nodes 236 nodes (471.24 TFLOPS)	CPU : Intel Xeon Gold 6126 (Skylake / 2.6 GHz 12 cores) 2 CPUs Memory : 192 GB
	GPU nodes 37 nodes (858.28 TFLOPS)	CPU : Intel Xeon Gold 6126 (Skylake / 2.6 GHz 12 cores) 2 CPUs GPU : NVIDIA Tesla P100 (NV-Link) 4 units Memory : 192 GB
	Xeon Phi nodes 44 nodes (117.14 TFLOPS)	CPU : Intel Xeon Phi 7210 (Knights Landing / 1.3 GHz 64 cores) 1 CPU Memory : 192 GB
	Large-scale shared-memory nodes 2 nodes (16.38 TFLOPS)	CPU : Intel Xeon Platinum8153 (Skylake / 2.0 GHz 16 cores) 8 CPUs Memory : 6 TB
Interconnect	InfiniBand EDR (100Gbps)	
Stroage	DDN EXAScaler (Lustre / 3.1PB)	

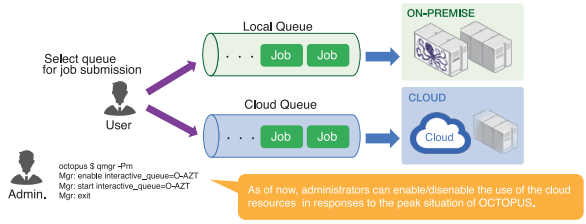
Taking it into consideration that we forward job requests onto OCTOPUS to Azure, virtual compute node should have more CPU cores and memory than OCTOPUS.

Job manager



IPMI wrapper and cloud controller have been developed so that

- We take advantage of NEC NQS II/JM's energy-saving functionality on the cloud.
- NQSII/JM can handle multiple cloud services simultaneously.



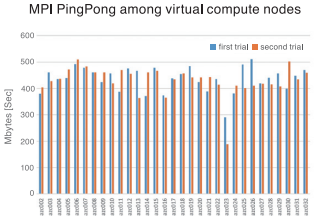
Ideally, these two queues should be integrated to a single queue so that jobs are processed on on-premise or the cloud in a user-transparent manner.

Evaluation

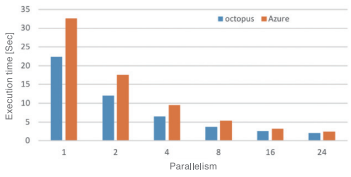
The following criteria have been used for evaluation of this OCTOPUS-Azure environment.

1. On-demand
   - Cloud resources should become available/unavailable in an on-demand way.
2. Transparency
   - Job submission to on-premise and cloud resources should not be different.
3. Selectivity
   - Users have to be able to specify whether they prefer the use of the cloud or not.
4. Equality
   - Computing results should be the "same" as in the cloud
5. High throughput
   - Throughput should be increased and then user waiting time should be reduced.

MPI PingPong among virtual compute nodes



GROMACS on OCTOPUS and Azure (single node)



Contact : sc19@ais.cmc.osaka-u.ac.jp http://www.hpc.cmc.osaka-u.ac.jp/en/