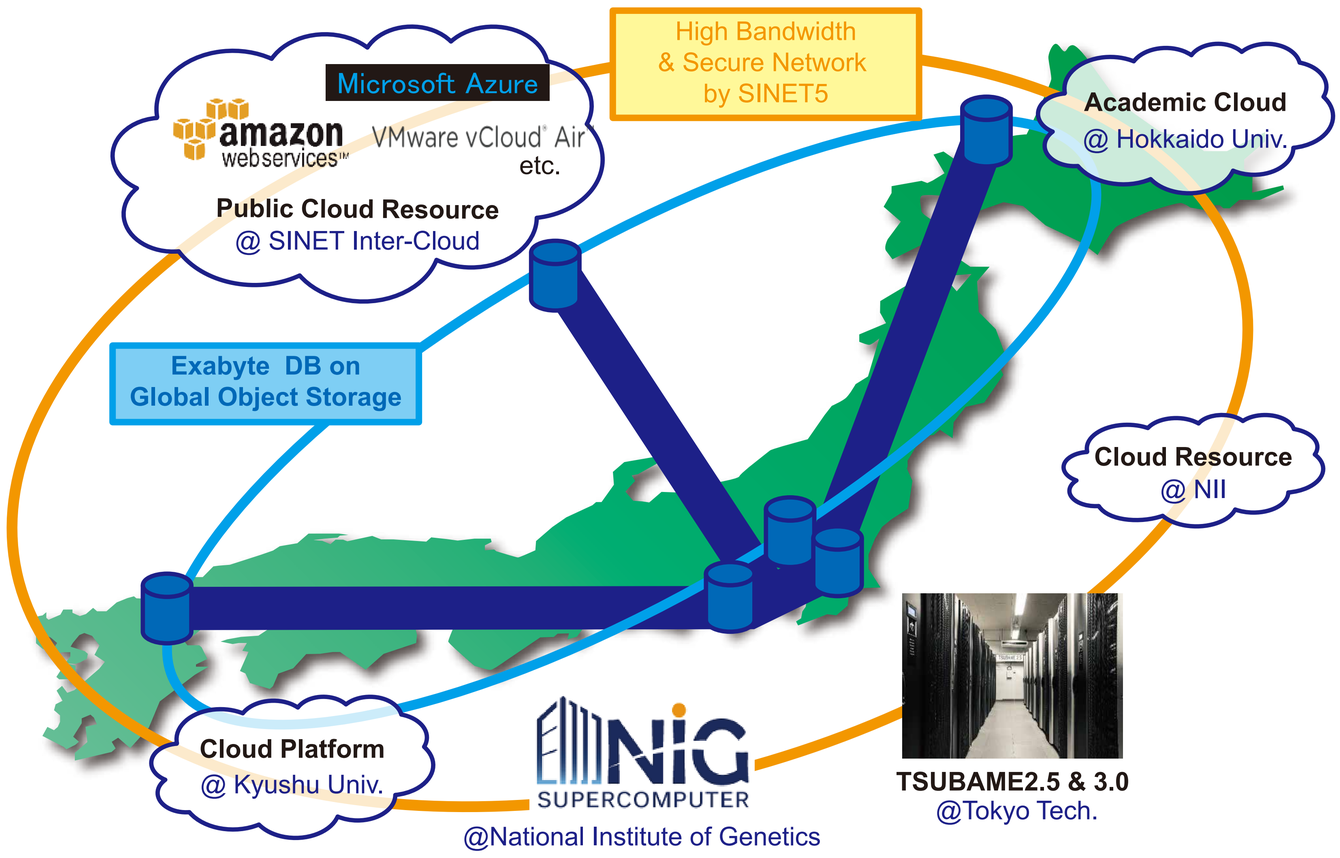
Big Data
Cloud Infrastructure for Big Data Analysis

Inter-Cloud Infrastructure for Big Data Analysis

Building a Testbed Infrastructure on Overlay Cloud

- Using SINET5 network infrastructure (100Gbps Network)
- Cooperation with various computing resources
  - Private cloud
  - Public cloud
  - Supercomputer
- Providing the Science Data Repository
  - Testbed: Peta Bytes class object storage
  - Real System: Exa Bytes class object storage



Collaboration work with NII, NIG, Hokkaido Univ. and Kyushu Univ.
For more details, Please visit following booths

#2908, "National Institute of Informatics"

#3658, "Hokkaido University"

#3637, "Kyushu University"

Acknowledgments. This research is supported by CREST, JST (Research Area: Advanced Core Technologies for Big Data Integration).

Cloud-Based Burst Buffer for I/O Acceleration

Background

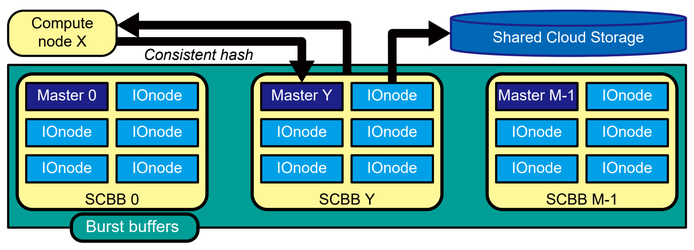
- Cloud storage cannot provide enough I/O throughput for data intensive applications.
- Loose consistency model in cloud storage
- No support of N-1 Concurrent write

Proposal

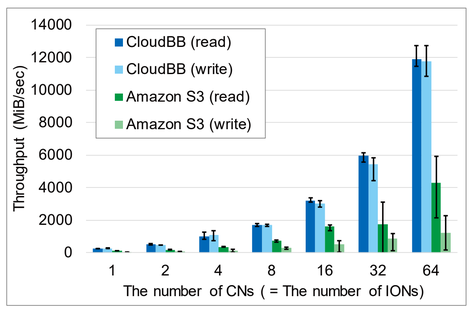
- Cloud-based Burst Buffer as a new tier in cloud storage hierarchy.
- Several dedicated instances as burst buffers to accelerate accesses by buffering intermediate data.

System

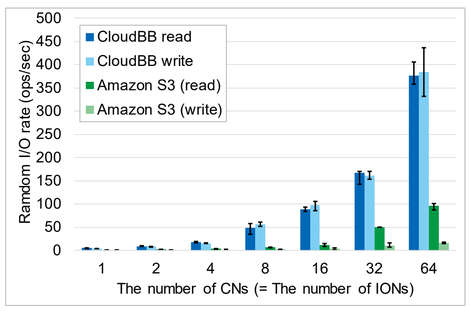
- The whole system consists of several Sub CloudBBs (SCBB)
- Each SCBB consists of a Master (metadata server) and several IOnodes (data server).
- Fault Tolerance supports



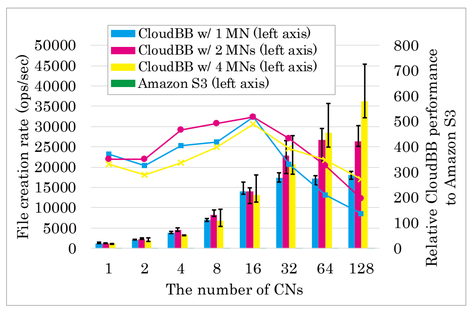
Sequential IO



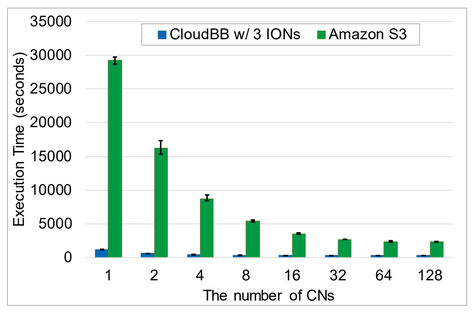
Random IO



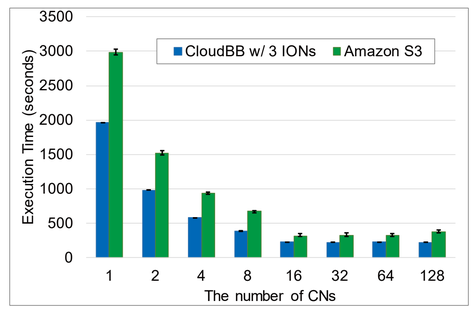
Metadata Performance



Real Application: Montage



Real Application: Supernovae



Acknowledgments. This research is supported by CREST, JST (Research Area: Advanced Core Technologies for Big Data Integration).

Cloud for HPC

Backgroud

Cloud platforms exhibit elasticity, flexibility, usability and scalability, which have been attracting users to use these environments as a cost effective measure to run their applications or businesses. However, the feasibility of running high performance computing applications on clouds has always been a concern mainly due to virtualization overheads and high-latency interconnection network.

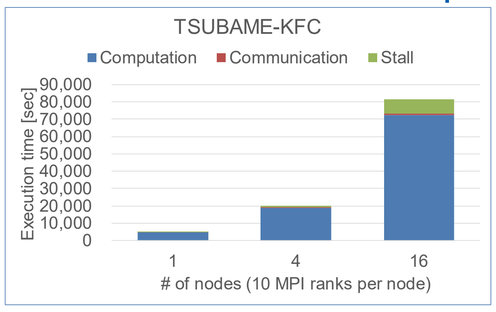
Goals

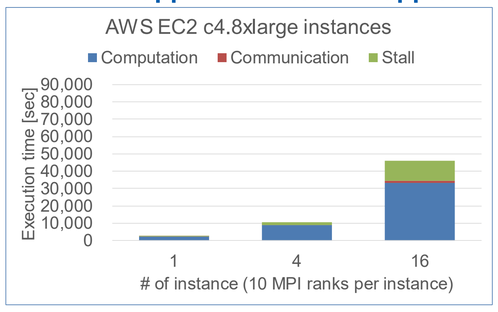
- To investigate the potential role of these virtual machines in addressing the needs of HPC and data-intensive workloads

Experimentation

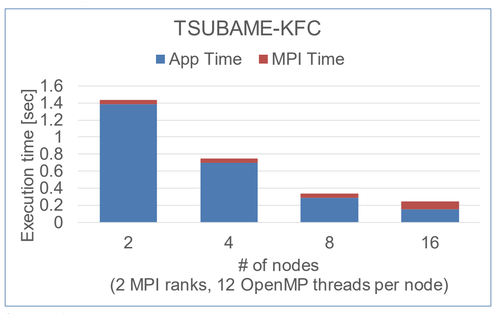
- Performance evaluation of applications on AWS C4 instances against the baseline results of a supercomputer, TSUABME-KFC

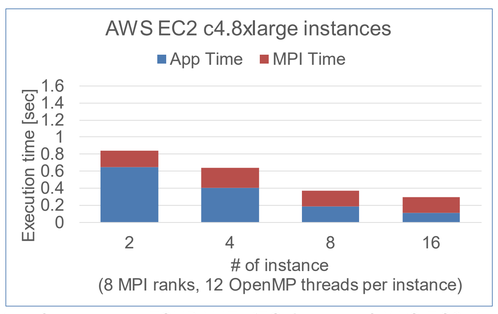
NICAM-DC-MINI: A compute-intensive miniapp from Fiber miniapp suite





Graph500 Benchmark at Scale 26





* 2 MPI ranks and 12 openmp threads per node achieved the best performance on TSUBAME-KFC. In case of AWS EC2, it was 8 MPI ranks and 12 openmp threads per instance.

TSUBAME-KFC: Intel Xeon E5-2620v2 x2, InfiniBand FDR
Amazon EC2 c4.8xlarge instance: Intel Xeon E5-2666v3, 10Gbps Ethernet

Poster at SC16:

"Towards Understanding HPC-Big Data Convergence Using Cloud Platforms", Shweta Salaria, Kevin Brown, Hideyuki Jitsumoto, Satoshi Matsuoka

Acknowledgments. This research is supported by CREST, JST (Research Area: Advanced Core Technologies for Big Data Integration).

http://www.gsic.titech.ac.jp/sc16