High Performance Computing Cloud - a PaaS Perspective

Pratima Dhuldhule, J. Lakshmi, S. K. Nandy

Supercomputer Education and Research Center
Indian Institute of Science, Bangalore

November 2, 2015

Overview

Introduction

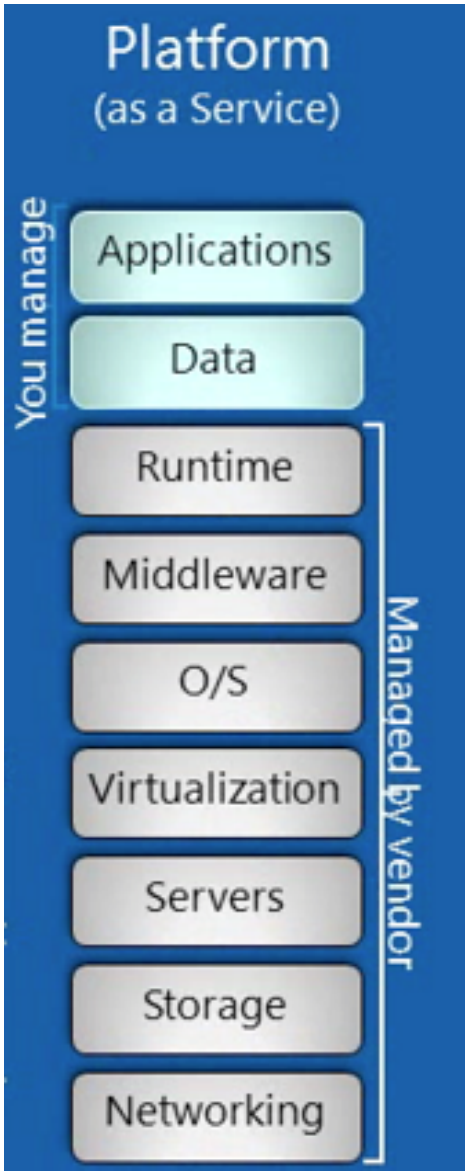
- Cloud computing is emerging as a latest compute technology
- Properties of cloud like availability, elasticity and flexibility have attracted attention of scientists for HPC workload
- HPC applications
  - dedicated resources
  - timely results
- Cloud platform
  - shared resources
  - virtualized environment

Motivation

- HPC applications
  - consume 90% of available resources
  - require low latency
  - timely results
  - tight interconnect network
- Disadvantages of HPC in Cloud
  - Shared resources - CPU jitter, I/O overhead, high latency
  - Security
  - Limited type of instances
  - No tightly coupled interconnect network

Motivation for PaaS Model

- Ability to provide on-demand platform
- Guarantees quality of performance
- Abstracts and controls underlying resources and gives choice of platform



Related Work

- HPC on cloud has been studied for long time
- [4] - attempts to improve performance of HPC application on cloud
- [1] - science clouds for scientific applications
- [3] - discusses strategies to completely remove hypervisor layer to overcome virtualization overhead

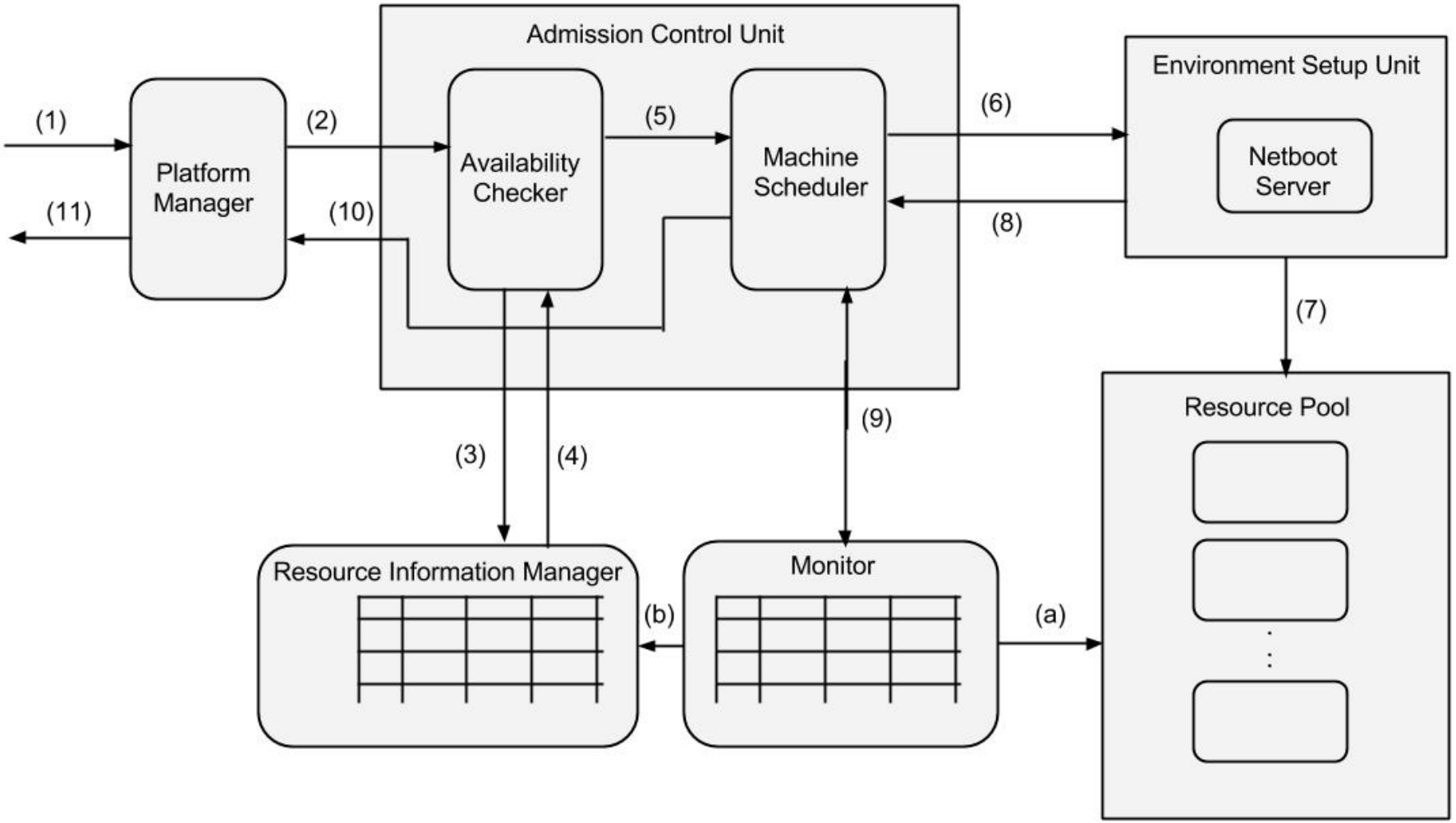
None of the above work talk about non-virtualized platforms for HPC applications.

Methodology

- An architecture to provide a platform with dedicated resources
- Build resource allocation strategies to meet HPC requirements
- Bring together PaaS advantages and HPC needs
- Design Goals
  - Isolated platform
  - Non-virtualized environment

HPC Cloud Computing

HPC Cloud Architecture



Case Study - Machine Failure

Failure can happen in two scenarios

- Failure during Netboot
  - ESU sends back error message
  - Another machine is made available
- Failure after allocation
  - Monitor detects failure
  - Scale up request is made

Sample XML request file

Sample XML request file

```
<?xml version="1.0"?>
<request>
    <hostrequirement>
        <processor>Pentium</processor>
        <cores>64</cores>
        <RAM>256</RAM>
        <disk>4048</disk>
    </hostrequirement>
    <environmentrequirement>
        <OS>linux</OS>
        <version>14.04</version>
        <library>mpi</library>
    </environmentrequirement>
</request>
```

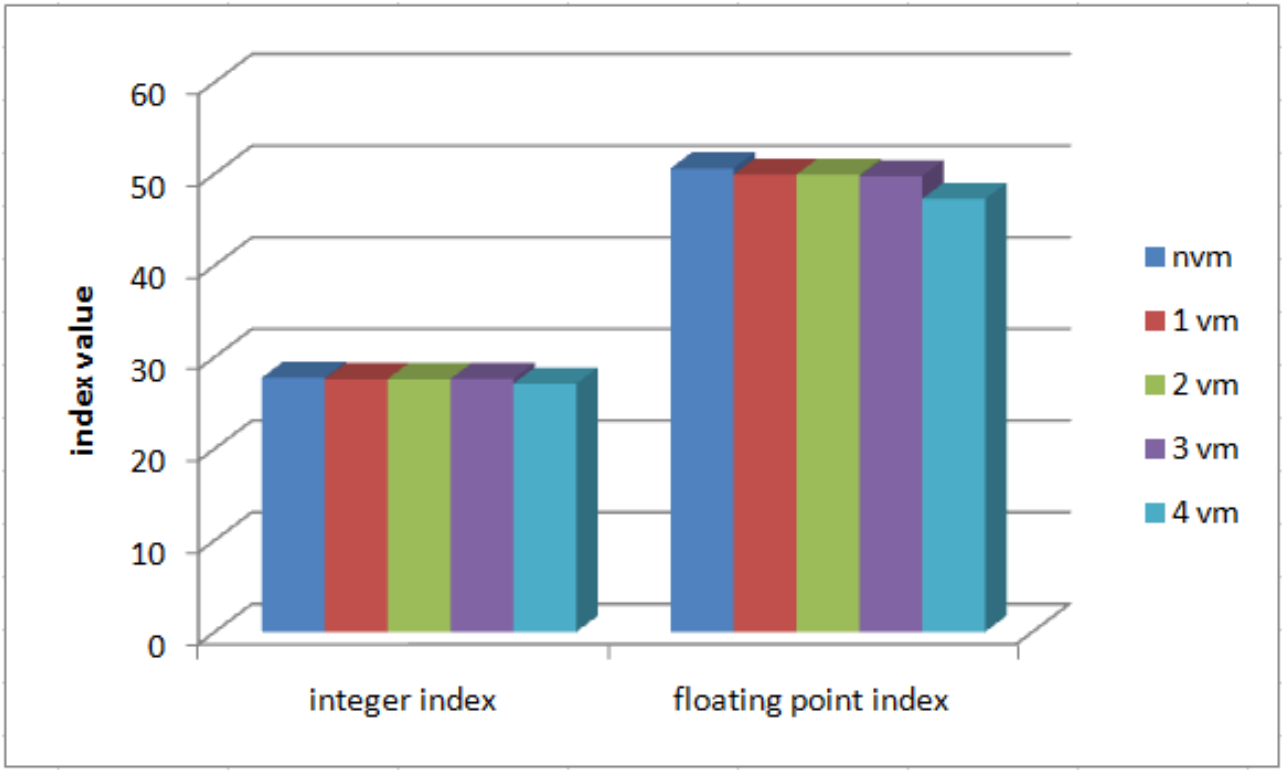
Experimental Setup

- Intel core i7 processor with 8 GB RAM and 1 TB SATA2 disk
- For virtual environment KVM was used as hypervisor
- native and virtual machine had following configuration
  - single CPU
  - 1 GB RAM

All VMs were running same workload simultaneously

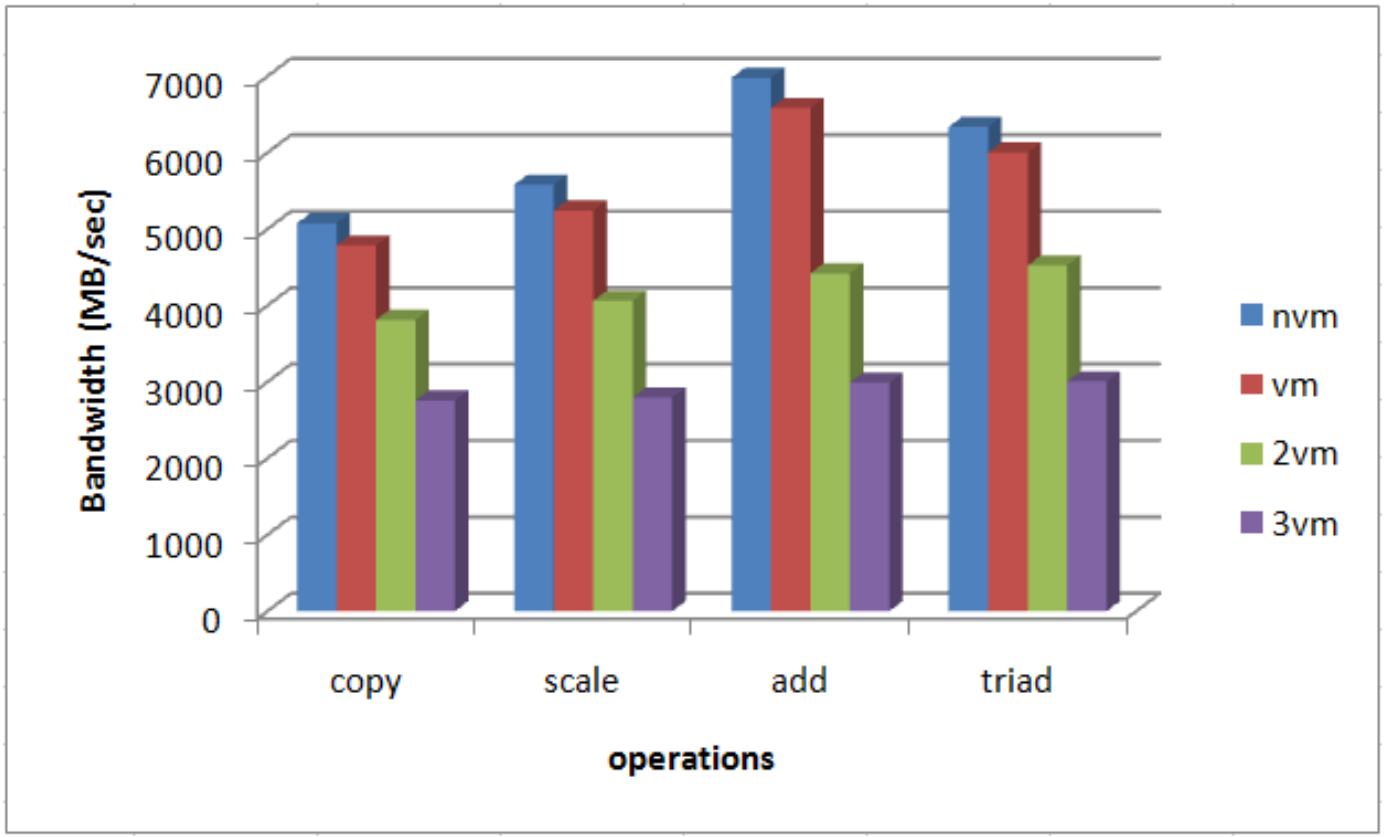
Results

Effect of virtualization on CPU



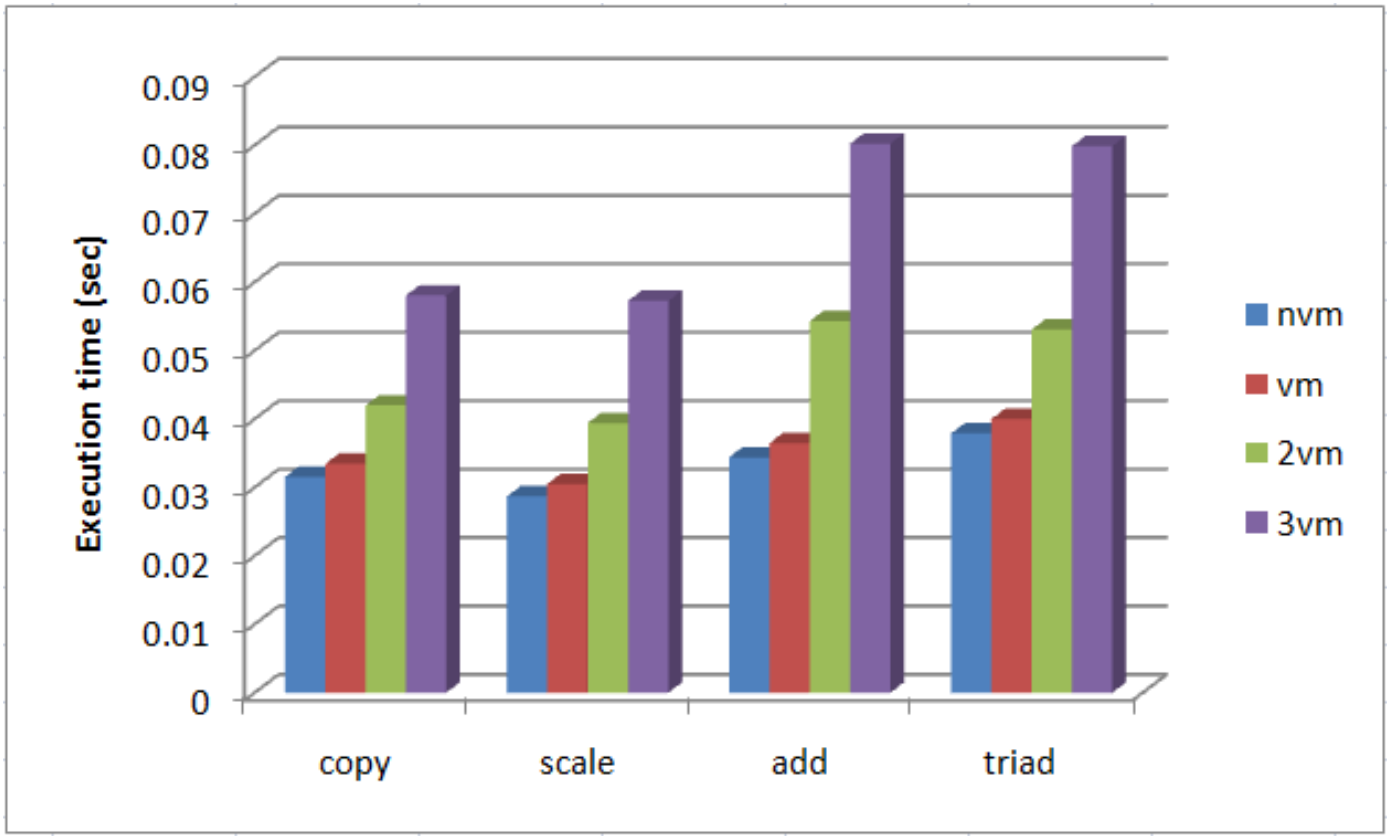
Results

Effect of virtualization on Memory Bandwidth



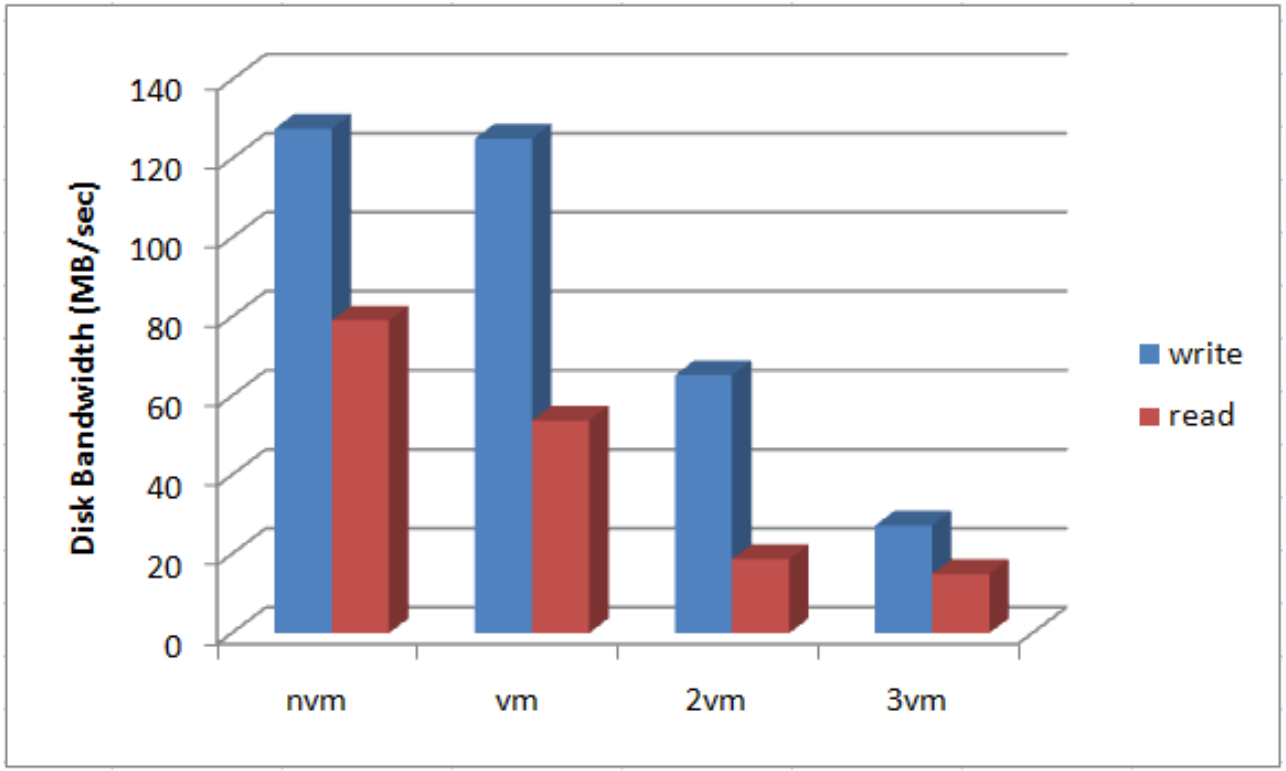
Results

Effect of virtualization on Memory Execution Time



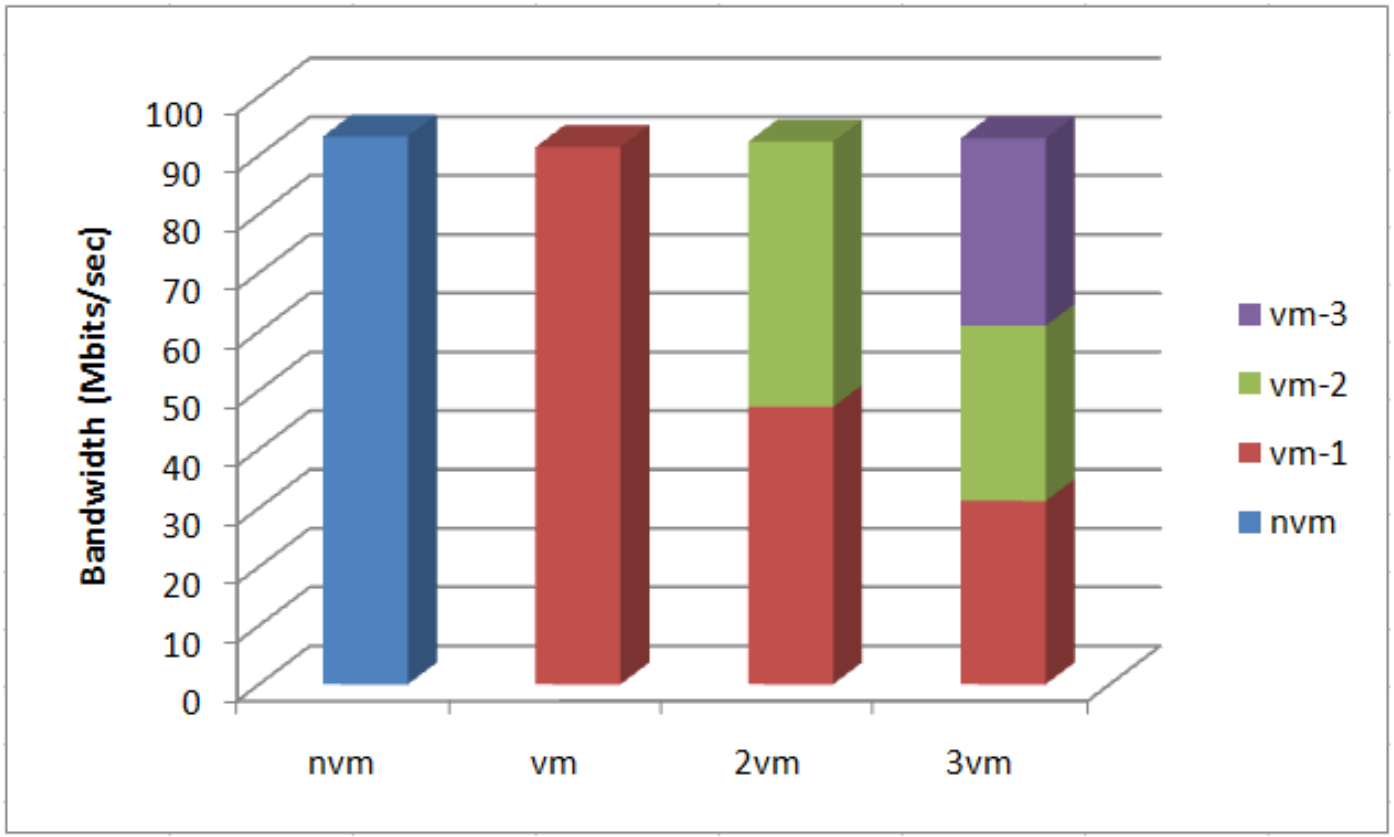
Results

Effect of virtualization on Disk IO



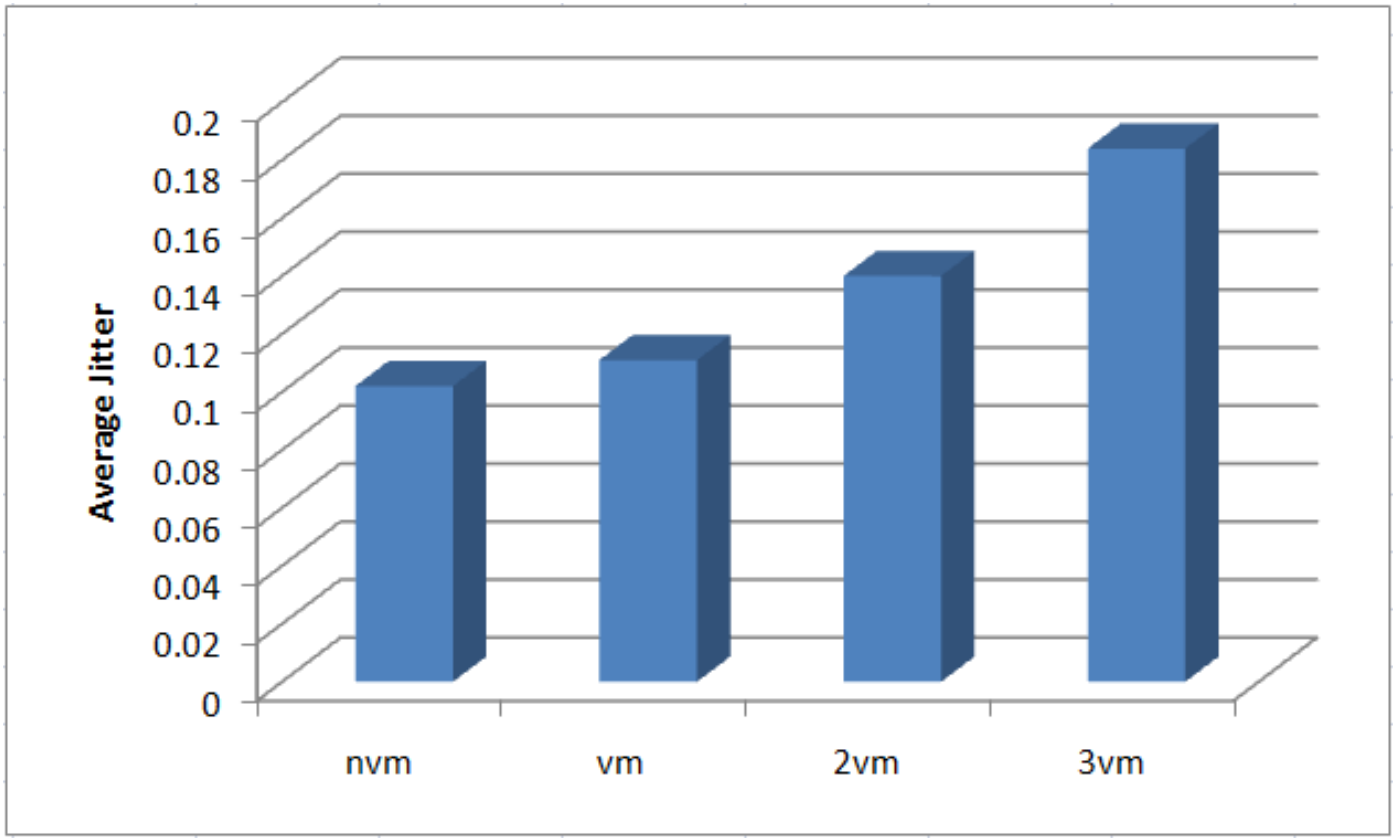
Results

Effect of virtualization on Networking Bandwidth for UDP



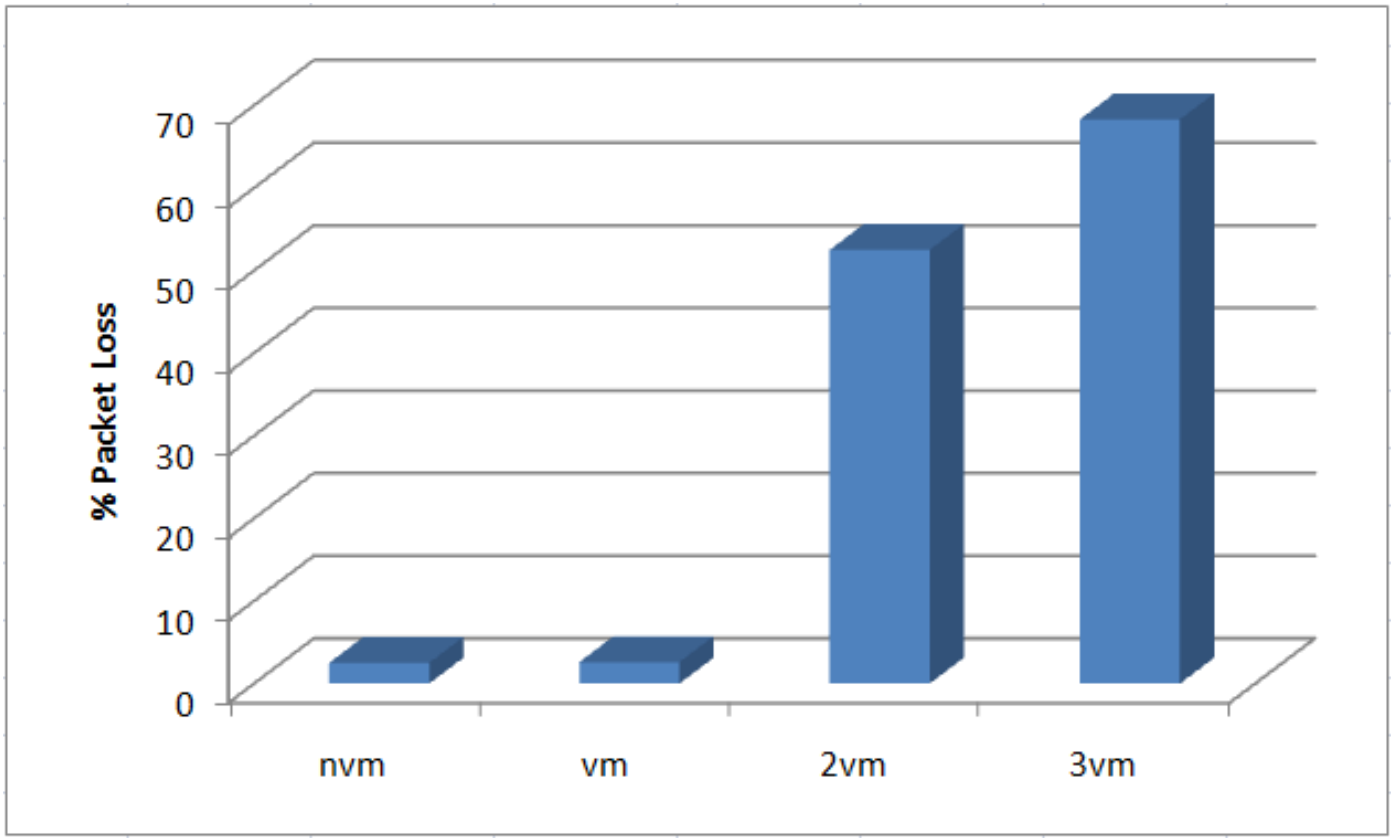
Results

Jitter for UDP



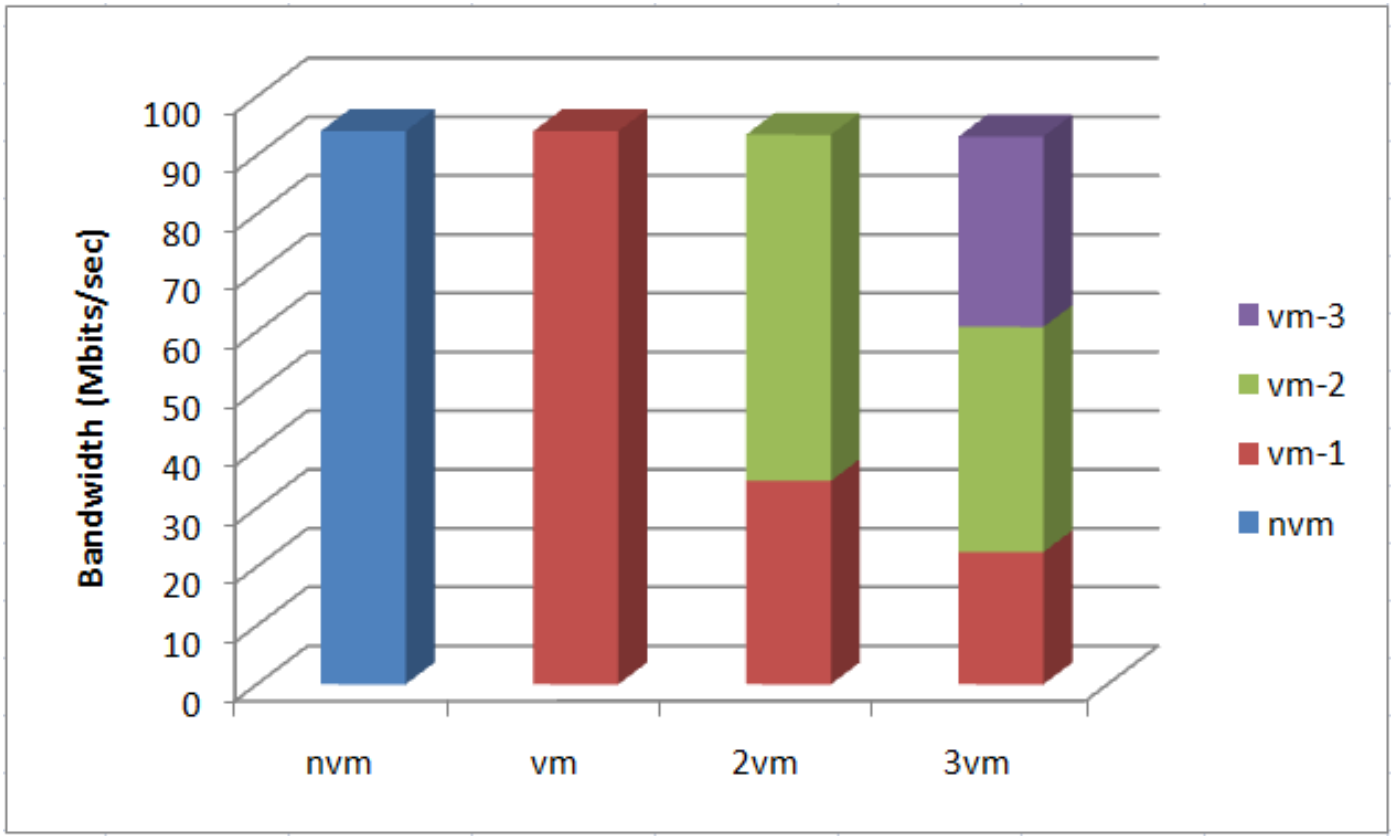
Results

Packet Loss for UDP



Results

Bandwidth for TCP



Conclusion

- In this work, we have designed an architecture for provisioning of isolated platform for HPC workload
- We evaluated performance across various physical resources and studied effect of virtualization on them
- We observed that virtualization causes performance degradation in HPC applications

Future Work

- Current architecture have features like starting, scaling and terminating a compute cluster
- Architecture needs further improved features such as pausing/resuming and saving state of compute cluster
- Testing of the architecture for real life HPC workload scenarios

References

[1] S. Srirama, O. Batrashev, and E. Vainikki, "Scicloud: scientific computing on the cloud", in Proceedings of the 2010 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing. IEEE Computer Society, 2010, pp. 579-580.

[2] S. K. Garg, C. S Yeo, A. Anandasivam, and R. Buyya, "Environment-conscious scheduling of hpc applications on distributed cloud-oriented data centers", Journel of Parallel and Distributed Computing, vol. 71, no. 6, pp. 732-749, 2011.

[3] E. Keller, J. Szefer, J. Rexford, and R. B. Lee, "Nohype: virtualized cloud infrastructure without the virtualization", in ACM SIGARCH Computer Architecture News, vol. 38, no. 3. ACM, 2010, pp. 350-361.

[4] A. J. Younge, J.P. Walters, J. Suh, D.-I. D. Kang, Y. Park, S. P. Crago, and G. C. Fox, "Towards a high performance virtualized iaas deployment".

Thank you