HPC Cloud: High Performance Computing in Infrastructure Clouds

Kate Keahey

keahey@mcs.anl.gov

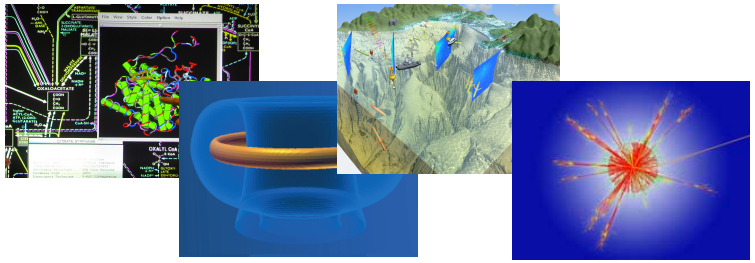
Argonne National Laboratory

Computation Institute, University of Chicago

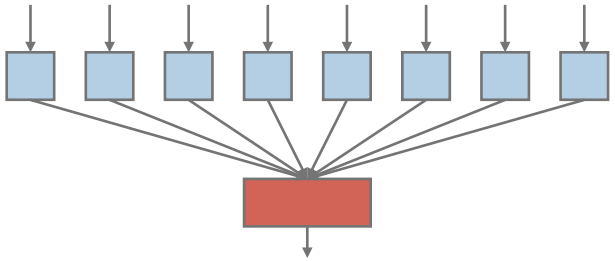
Clouds of All Types

Software-as-a-Service (SaaS)

Community-specific tools, applications and portals



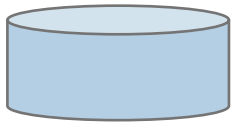
Platform-as-a-Service (PaaS)

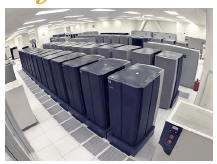


Microsoft

Google

Infrastructure-as-a-Service (IaaS)





amazon web services™

GOGRID beta
A ServePath Company

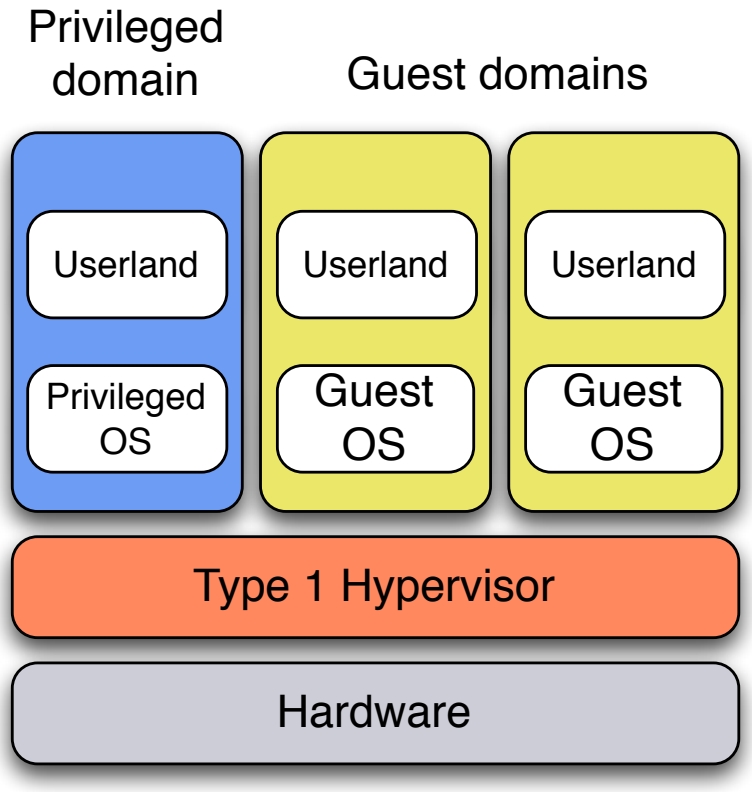
Control

Specialization

Technology Trends

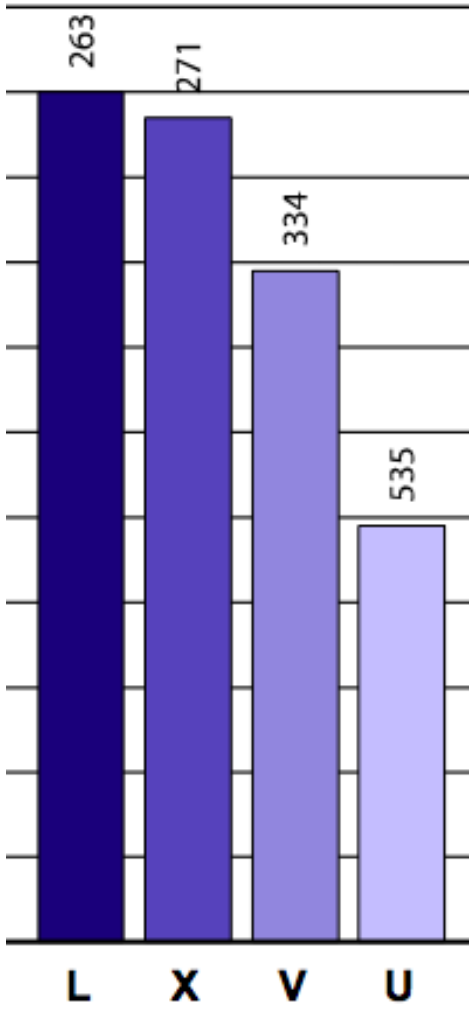
Virtualization

- System-level virtualization
  - Emulates a computer similar to a real one
  - The VM runs a full OS
- VM images
  - Snapshots
  - Migration
- Scheduling



Widespread Virtualization

- From IBM in 1967…
- …to Xen in 2003
  - The Need for Speed
  - Open Source
- Other hypervisors
  - KVM
  - Palacios, KHVMM



From Pratt et al., SOSP 2003

Virtualization: Game-changing Benefits

- Benefits to users
  - Control over environment
    - Touches all aspects from convenience to privilege
  - Convenient packaging/distribution mechanism
    - No makefiles, no validation and revalidation, facilitates sharing
- Benefits to providers
  - Security implications of isolation
  - Speed of deployment, switching between environments (critical difference from reimaging)
- Migration and snapshotting

Infrastructure Clouds: Game-Changing Benefits

- Benefits to users
  - On-demand access: manage peaks in demand
  - Pay-as-you-go: manage “valleys” in demand
  - Viable infrastructure outsourcing model
- Benefits to providers
  - Consolidation and economies of scale
- Retail versus wholesale resources

Clouds and HPC

Clouds and HPC

Can HPC workloads be run in a cloud?

HPC Cloud

Can a supercomputer operate as a cloud?

The Nimbus Project @ ANL

High-quality, extensible, customizable,
open source implementation

Nimbus Platform

Context Broker | Cloudinit.d | Elastic Scaling Tools

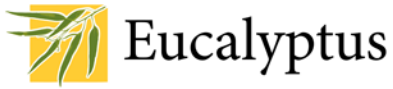
Enable users to use IaaS clouds

Nimbus Infrastructure

Workspace Service | Cumulus

Enable providers to build IaaS clouds

Enable developers to extend, experiment and customize



The Red Question (RQ)

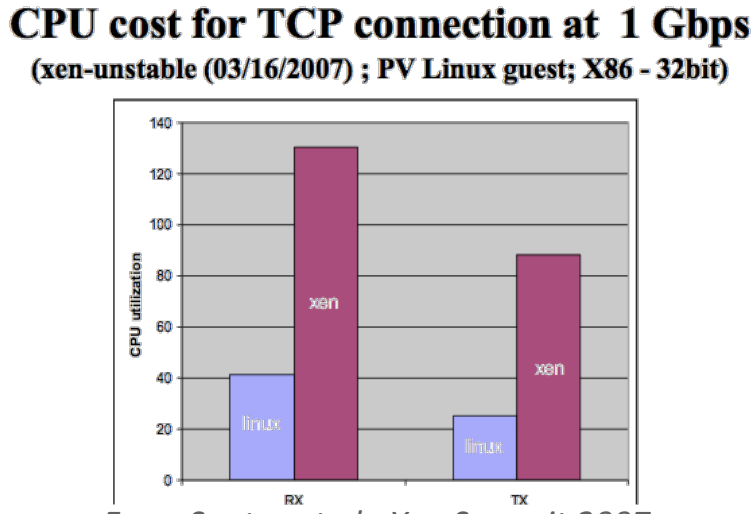
Can HPC workloads be run in a cloud?

Napper et al., "Will Cloud Computing Reach Top500?" (The answer was: "No")	**2009**
Virtual Cluster from AWS #146 on Top500 (currently #42)	**2010**

Total time to change your mind: **1 year**

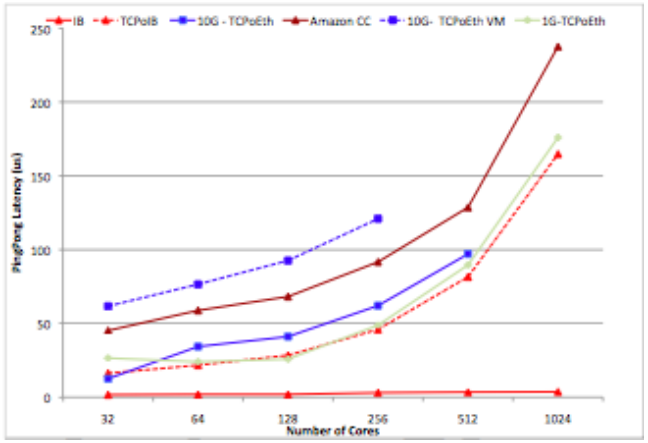
RQ: Performance

- I/O Performance
  - Bandwidth
  - Latency
- Instability



From Santos et al., Xen Summit 2007

How good are clouds for HPC?
Resources:
- The Magellan Report
- "Comparisons: not as Odious as Once Thought", see www.sciencecloud.org/blog



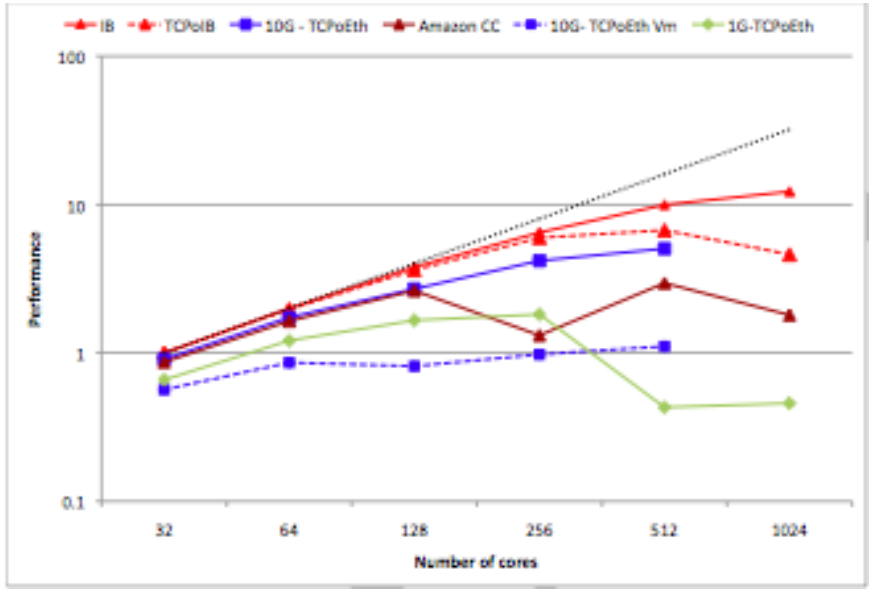
From Ramakrishnan et al., PMBS'11

RQ: Storage

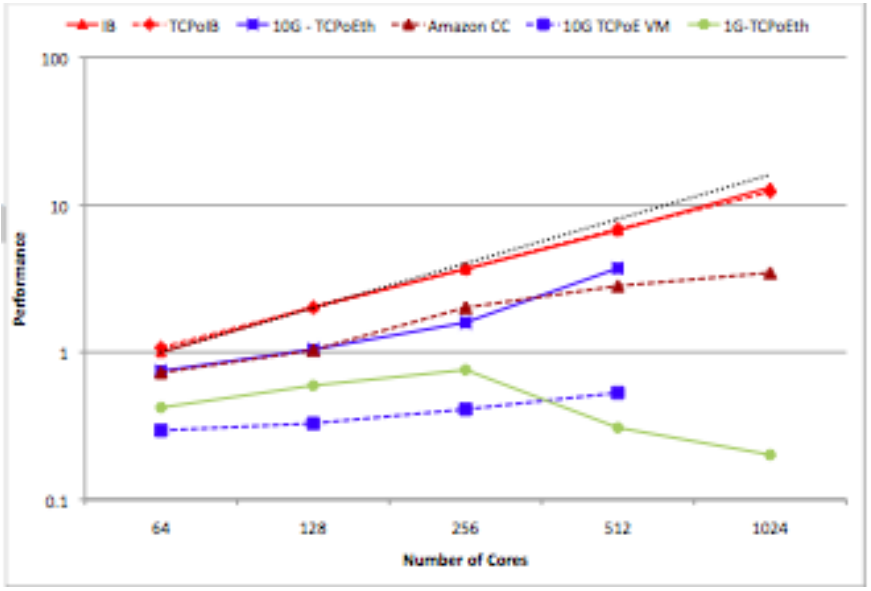
- Performance challenges
  - I/O impact
  - Buffer caching not enabled
  - Bandwidth contention
  - Variability
- Getting the custom infrastructure
  - Storage options of all shapes and sizes
  - How do we combine offerings?
- Price performance considerations
  - Instance service levels versus HPC needs

RQ: Overall Performance

Overall impact on tightly-coupled applications



PARATEC

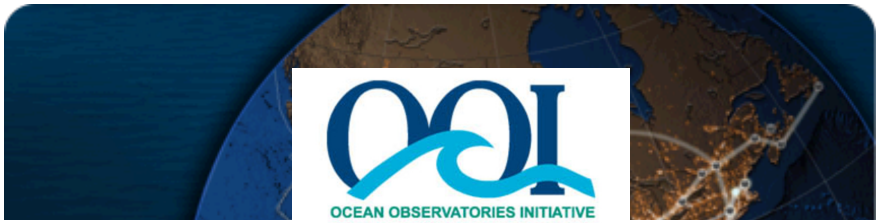


MILC

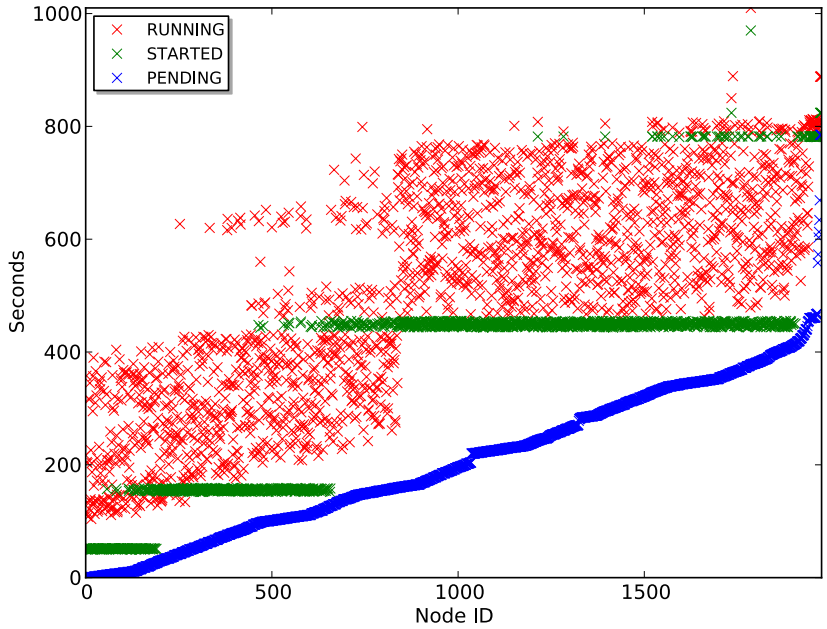
From Ramakrishnan et al., PMBS'11
Also see Jackson et al. , CloudCom'10

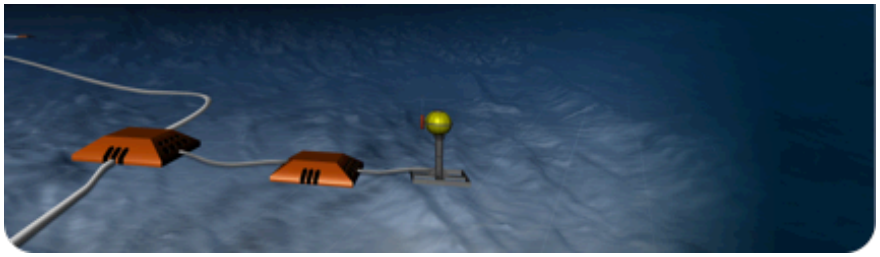
RQ: Size and Scale

- "Utility Supercomputing"
- Record (CycleComputing)
  - Computational chemistry app
  - ~6,000 instances
  - ~50,000 cores
  - ~$5,000 per hour
- Nimbus for OOI
  - BigData analytics
  - Elastic scaling and HA
  - ~2,000 instances



Lower bound deployment time by node number





RQ: "Other Considerations"

- Special hardware
  - E.g., Amazon GPUs instances
- Noise
  - Significant variability
- MTBF and failure rates
  - Both a hardware and software consideration
  - Significant lack of reliability
  - Ask Franck…

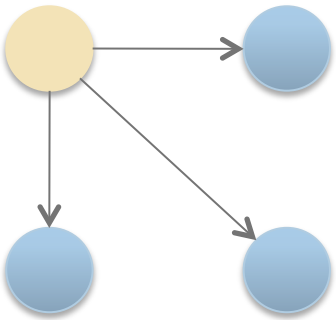
See Jackson et al. , CloudCom'10

How good are clouds for HPC?
"Mohammad and the Mountain"
see www.sciencecloud.org/blog

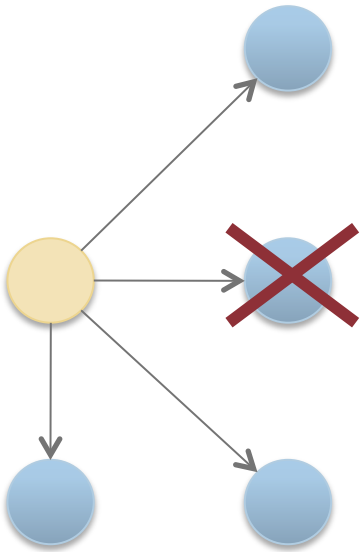
Programming Models

- Leverage on-demand, large provider pool
- Adapt to failures
  - Master/Slave example

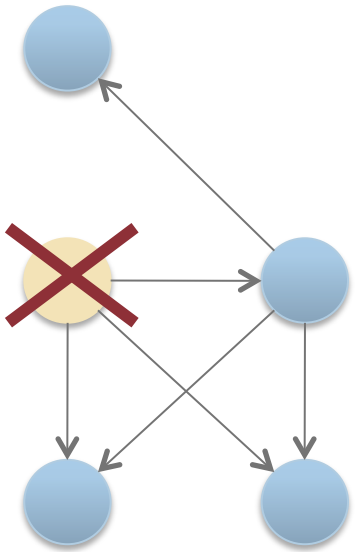
Auto-scale



Redeploy



Role transfer + leader election



RQ: Open Issues

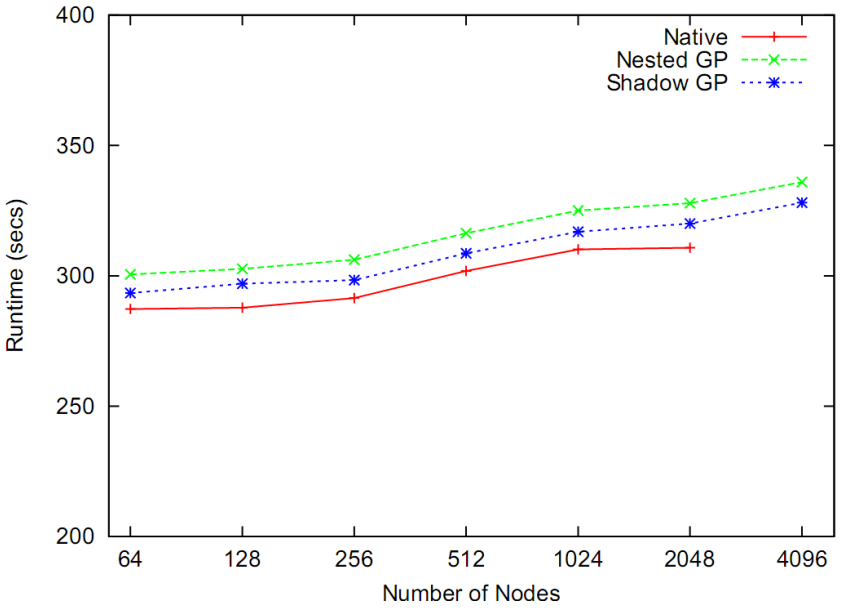
- More understanding of performance
  - Still need understanding of external connectivity, storage/ performance tradeoffs, MTBF & noise
  - Ongoing issue as the offerings change dynamically
- Performance characterization
  - Lightweight, easy to run, as conclusive as possible
- Storage space/options still largely unexplored
- Programming models

The Blue Question (BQ)

- Is it feasible to turn a supercomputer into a cloud?
- **Challenges:**
  - Hypervisor challenges
  - Deployment speed
  - Resource management model

BQ: Hypervisor Issues

- **Challenges:** architecture, I/O drivers, and performance
- New hypervisors
  - KHVMM (Blue Gene/P)
- Xen-IB (since ~2006)
- "Symbiotic virtualization"
  - Passthrough I/O
  - Preemption control
  - Optimized paging
- Research software

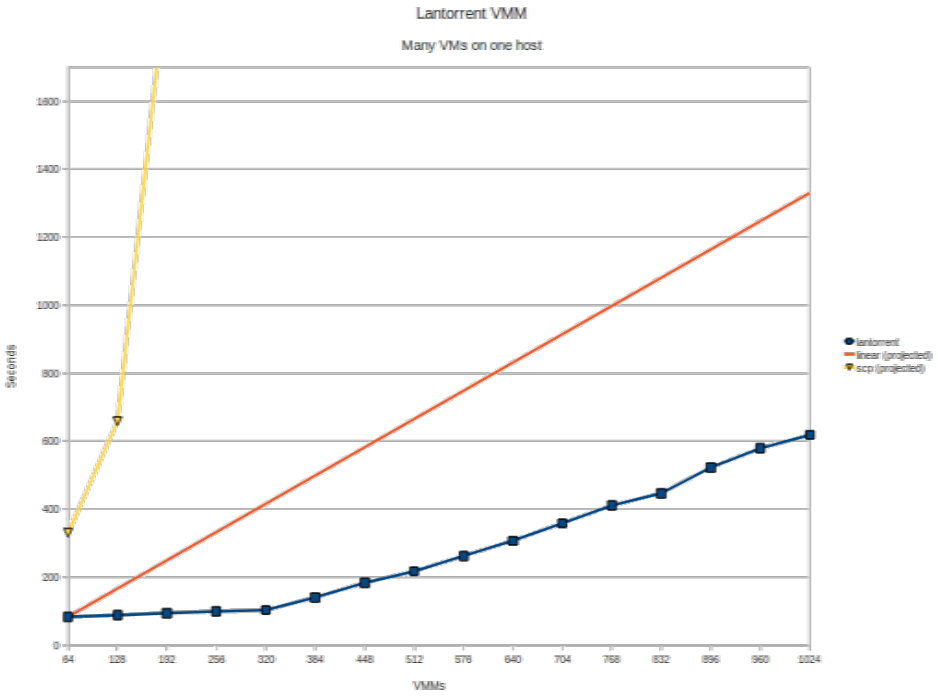


CTH (shockwave simulation)
within ~5% of native performance
On Cray XT4 (Sandia)

From Lange et al., VEE' 11

BQ: Deployment Scale and Speed

- Moving images is the main component of VM deployment
- **Challenge:** make image deployment faster
- LANTorrent: the BitTorrent principle on a LAN
  - Streaming
  - Minimizes congestion at the switch
- Detecting and eliminating duplicate transfers
- **Bottom line**: a thousand VMs in 10 minutes



Evaluation using the Magellan resource At Argonne National Laboratory

- Other approaches:
  - *Nicolae et al., HPDC'11*
  - *Riteau et al., QCOW*

BQ: Resource Management

- **Challenge:** high performance versus cost
- Multi-tenancy
  - Noisy neighbor
  - Interleaving needs
  - Not suitable for HPC with current methods
- Single-tenancy
  - Utilization challenge
  - Preemptible instances and spot instances: increase utilization without sacrificing the ability to respond to on-demand requests
  - Preemptible with snapshotting: increase utilization without sacrificing the ability to resume computation

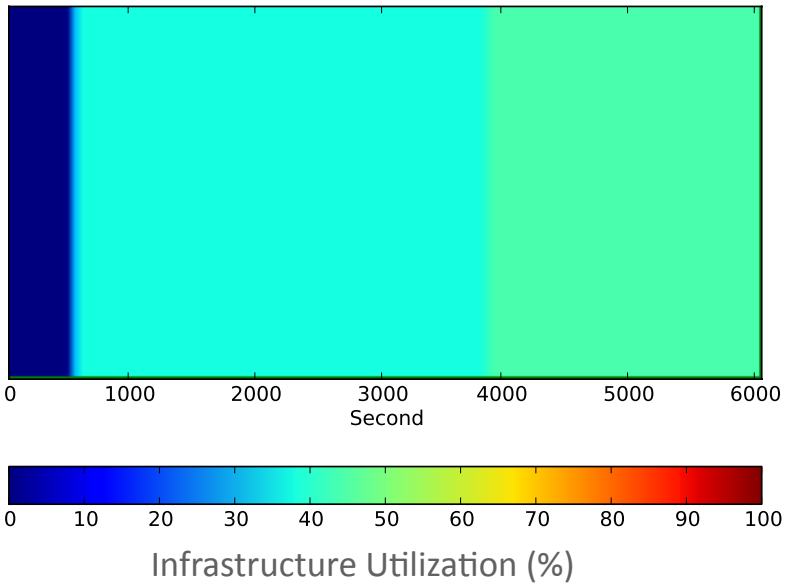
  Paper: Sotomayor et al., HPDC'08

BQ: Resource Management (2)

Preemption Disabled

Average utilization: 36.36%

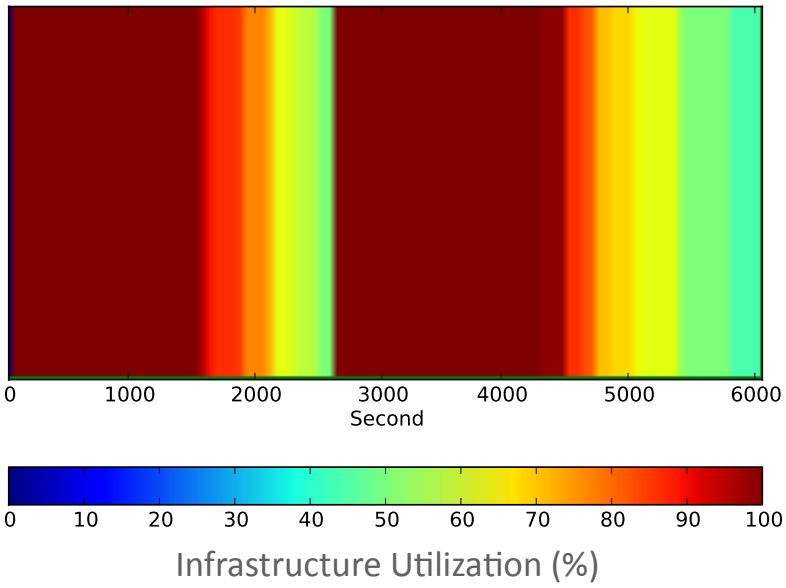
Maximum utilization: 43.75%



Preemption Enabled

Average utilization: 83.82%

Maximum utilization: **100%**



From Marshall et al., CCGrid'11

BQ: Open Questions

- Hypervisors
  - Rapid progress but still many tradeoffs to examine
  - Need robust, open source solutions
- Resource management
  - Combine various concerns to allow users to build better virtual clusters
  - Utilization, policies, energy saving, etc.
- Storage management
  - Service levels
  - Combining storage and compute clouds
- Image management

Why Do We Care?

- Makeup of Top500 by Franck:
  - 87% commodity clusters (similar to Amazon CC)
  - 13% luxury computing
- Cloud computing creates a “critical mass”
  - Attracts applications and investment
- Can we afford to not support it?

Open Issues

- Performance characterization
  - Conclusive, easy to run, and lightweight
  - Reflecting both application requirements and platform idiosyncrasies
  - “Cloud500”
- Filling the gap between theory and practice
- “Shining star” demo
- Models
- Opportunities: big data

Build-a-Collaboration

- KerData team (Rennes)
- Myriads team (Rennes)
- Avalon team (Lyon)
- LBNL
- Northwestern
- University of Colorado
- ISI
- OpenCirrus (international)