Integrating Cloud & Cyberinfrastructure

Manish Parashar

NSF Cloud and Autonomic Computing Center
Rutgers, The State University of New Jersey
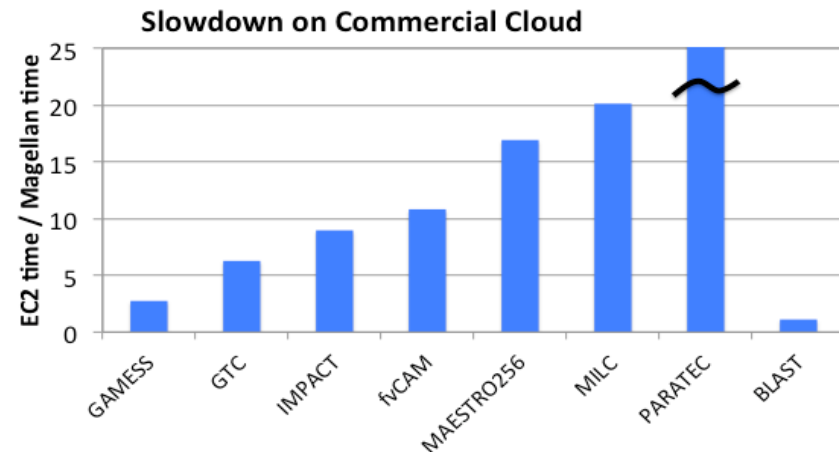
Clouds and Science

- Clouds are rapidly joining traditional CI as viable platforms for scientific exploration and discovery
  - What application formulations and usage modes that are meaningful in such a hybrid infrastructure?
  - How can application workflows effectively utilize it?

- Possible usage modes:

  - Managing complexity by outsourcing the mundane

  - Democratization

  - Application driven by the science, not available resources!

    - Cloud abstractions for science (beyond IaaS/PaaS/SaaS)?

- Many challenges

  - Application types and capabilities that can be supported by clouds?

  - Can the addition of clouds enable scientific applications and usage modes that are not possible otherwise?

  - What abstractions and systems are essential to support these advanced applications on different hybrid platforms?

HPC in the Cloud

- Run applications on commodity clouds (e.g.,EC2)
  - Quick start-up, easy to use
  - Good performance for specific application classes
- However not suited for many applications
- And costs can add up quickly……

- For example, at NERSC*
  - Baseline Cloud is 8x more expensive
  - Effective performance of the cloud is a minimum of 16x less cost effective (and worst case 320x less effective) for DOE midrange HPC workloads

Slowdown on Commercial Cloud

EC2 time / Magellan time — GAMESS, GTC, IMPACT, fvCAM, MAESTRO256, MILC, PARATEC, BLAST

* HPC in the Clouds, John Shalf Feb 3, 2011

CS&E on the Cloud – Moving beyond the obvious candidates

- New application formulations
  - Asynchronous, resilient, elastic
  - E.g., Asynchronous Replica Exchange Molecular Dynamics, Asynchronous Iterations

- New delivery models
  - Client + Cloud accelerators, HPC-as-a-Cloud
  - E.g., Matlab/Excel + EC2

- New hybrid usage modes
  - Cloud + HPC + Grid + …

Integrating Cloud & CI

- **HPC in the Cloud**, where researchers outsource entire applications to current public and/or private Cloud platforms.

- **HPC plus Cloud**, focused on exploring scenarios where Clouds can complement HPC/Grid resources with Cloud services to support science and engineering application workflows, for example, to support heterogeneous requirements, unexpected spikes in demand, etc.

- **HPC as a Cloud**, focused on exposing HPC/Grid resources using elastic on-demand Cloud abstractions, aiming to combine the flexibility of Cloud models with the performance of HPC systems.
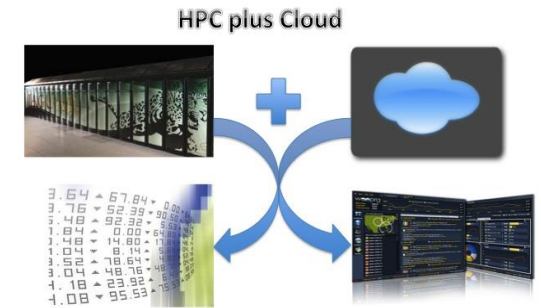
Integrating Cloud & CI

- ❏ Overview of the session
  - ■ Opening/context – Manish Parashar, Rutgers
  - ■ XSEDE perspective - John Towns, NCSA
  - ■ FutureGrid perceptive - Geoffrey Fox, Indiana Univ.
  - ■ OCC perceptive - Robert Grossman, Univ. of Chicago
  - ■ Magellan experiences - Shane Canon, LBNL
  - ■ GENI perspective - Chip Elliott, BBN

- ❏ Key questions
  - ■ What are the new opportunities that will be provided by integrating Cloud and Cyberinfrastructure?
  - ■ What would be the key challenges, technical as well as from a users perspective?
  - ■ What are immediate next steps, for agencies and for the community?