XSEDE and Cloud Technologies: status update

John Towns
PI and Project Director, XSEDE
Director, Distributed Cyberinfrastructure Programs, NCSA
jtowns@ncsa.illinois.edu
Developing XSEDE’s Strategy for Cloud Technologies

• Recognize the following:
  – cloud technologies likely have a role to play in the XSEDE environment
  – we (XSEDE and friends) don’t know all there is to know
  – there is quite a bit already known in the community, but unlikely covering what we need to know
  – need to figure out what we know and what we need to know

• Parameters of the exercise:
  – only looking to incorporate capabilities that are complimentary to what exists in XSEDE environment already
    • what could you do with cloud that you can't do on XSEDE otherwise
  – leverage what is known and identify what we need to know
    • cast as specific questions to be answered
  – develop any necessary pilot to answer specific questions
Current Activity: Identifying what we know via Use Cases

• Wealth of information that is out there, but not well collected/organized
  – various experiments have been done
  – what are cloud characteristics that are important for your use case?

• Documenting known use cases and what has already been learned
  – FutureGrid
  – work done by Manish Parashar
  – Microsoft Azure experiences
  – still to engage Amazon Web Services, Google, other projects…

• Initial publically available document coming in the next couple of weeks
  – initial discovery and common use cases that make sense for XSEDE
Some Current Example Use Cases (1)

• **Science Gateways**: domain specific web portals
  – provide the particular community of researchers access to the common resources/services
  – also can provide entrée’s into more traditional HPC environments
  – Examples:
    • HubZero: [http://hubzero.org](http://hubzero.org), [http://hubzero.org/sites](http://hubzero.org/sites)
    • XSEDE Science Gateways: [https://portal.xsede.org/science-gateways](https://portal.xsede.org/science-gateways)

• **Collaboration**
  – team wiki’s & web sites for communication, coordination, planning, documentation and document/data sharing
  – Examples:
    • Citrus Greening-HLB Genome Resources (CG-HLB): [http://citrusgreening.org/index.html](http://citrusgreening.org/index.html)
Some Current Example Use Cases (2)

- **Domain Specific Computing Environments**
  - custom software environments for data analysis/pre & post processing stages of scientific workflows or Event Driven Science
  - virtual operating systems/application software used remotely via ssh and/or xterms
  - Examples:
    - EnKF based history-matching workflow for oil reservoir modeling applications using Ranger (TACC), Clouds (EC2) and FutureGrid.
    - replica exchange Molecular Dynamics application reformulated for heterogeneous platforms integrating Ranger (TACC), OSG and Clouds (EC2)
    - Monte-Carlo value-at-risk-computations running on OSG and Clouds (EC2)

- **Burst Resources**
  - additional resources on demand to augment local resources
  - Examples:
    - EnKF based history-matching workflow on XSEDE resources (Ranger) complemented with Amazon EC2 Cloud instances to achieve user objectives
    - Use of clouds and autonomic cloud-bursting to support a medical image registration enabling integration of local computational environments and cloud services on-the-fly
Thoughts on Potential Further Info Gathering

• Want to open up ability for many to submit actual/potential use cases

• Propose to develop an information gathering activity
  – create a data collection web space
  – allow folks to add things
    • encourage Microsoft, Amazon, Google, FutureGrid, etc to have their users add use cases
  – collect concrete data on what works and what does not work
Developing a Plan for Pilots

• Two broad categories of pilots
  – use cases already successfully using clouds
    • Likely no pilot really needed; glean knowledge from experiences
  – use cases with potential for important capabilities
    • need a clear plan for executing a pilot

• Pilot Outcomes/Evaluation
  – describe research that can be supported
  – describe cloud resources required
  – define metrics of success
  – describe XSEDE integration requirements
A Couple of Key Integration Issues

• Authentication, Authorization, Accounting
  – How do we bridge XSEDE authentication, authorization, and accounting infrastructure with a cloud service provider?
  – What are peculiarities with integration with commercial vs. academic providers?

• Chargeback for commercial providers
  – How do we integrate commercial providers fully into portfolio of resources in a way that naturally allows for chargeback of commercially provided resources?
Our reach will forever exceed our grasp, but, in stretching our horizon, we forever improve our world.