Infrastructure Clouds MAGIC Meeting

Chris Hill, MIT September 7th 2011

interested in modeling of Earth and planetary systems to both understand basic physics and monitor Earth.

Main digital tools are observational datasets and numerical models in Fortran, C plus python, Octave etc... analysis

Question responses

- Describe how your community uses infrastructure clouds - I

    – Some examples

Many Task Computing for Multidisciplinary Ocean Sciences: Real-Time Uncertainty Prediction and Data Assimilation

Evangelinos, Lermusiaux, Xu, Haley, Hill
[IEEE Trans. Parallel Distrib. Syst. 22(6): 1012-1024 (2011)](#)

Cloud Computing for parallel Scientific HPC Applications: Feasibility of running Coupled Atmosphere-Ocean Climate Models on Amazon's EC2.

Evangelinos, Hill
CCA 2008.

Many task example – scenario where compute resources need to scale up to coincide with field experiments. Problem is a model-data synthesis for real-time ocean state estimation. Field experiments occur over few week periods each year and require ensembles of models run in real-time. Bursty demand more cost effective to sequester resources as needed. EC2 IC worked v. well for this.

 CCA example – complex coupled model with MPI, but relatively modest scale 10-100 cores. EC2 IC provides way to package system in a box for non-expert users, including GUI interface etc… Useful for education as well as research applications.

Question responses

- Describe how your community uses infrastructure clouds - II
  - Some examples
    - Cloud Infrastructure Technology for Education

      .pdf as interface!

      hyperlinks launch browser forwarded to on demand applications that allow simulation, viz etc... connected to .pdf content. Use EC2 IC to provide back end on demand.

      works from tablet, smart phone (sort of) etc...

      test deployments in two MIT classes this fall. Tests in Mass public schools next spring (CI-Team with UMass CAITE).

Question responses

- - What drew you to infrastructure clouds, i.e., what are the benefits for your community of using infrastructure clouds in order of significance?
  - Environment for packaging complex application as a virtual artifact.
  - Simple interface between hardware and packager, with a lot of flexibility.
  - Can have VM per simulation configuration, customized to appropriate needs.

Question responses

- - What are the challenges of using infrastructure clouds for your community, i.e., what makes them difficult to use in order of significance?
  - Performance is sub HPC cluster
  - Persistent storage is particularly pricey
  - Lack of a standard high-level interface to the process of subscribing, launching etc...
  - Work best for providing non-expert users access to research tools. Implies including interface etc.. in VM.

Question responses

- - How do infrastructure clouds compare to other options for outsourcing computation from the perspective of your community's needs?
    - Certainly IC are not a replacement to traditional HPC centers.
    - However, IC provides a lot more flexibility. For example in a class I teach on GPU computing it I created an image with example codes and full software stack that a student can use on EC2 with GPU. This is a lot easier for me than using a local cluster or using an HPC resource like Keeneland. On other hand for large GPU computations supporting some hydrology research I have the NSF Keeneland system is a great resource.