**Science Clouds and Campus Clouds**

Geoffrey Fox

Indiana University

**Abstract**

Cloud computing offers several interesting features for academic users for both large research applications and the myriad of modest jobs typical of large campus systems. Some of the attractions of clouds are shared with those for commercial clouds; namely cost effectiveness and ease of dynamic scaling. Further as well as cloud infrastructure, cloud-related technologies like Hadoop could be important in supporting the growing importance of data intensive science in particle physics, biology, earth/environmental science among other disciplines. Currently clouds seem particularly attractive in supporting “workflows of pleasing parallel jobs” that are common in these areas. However current clouds are not well suited to applications that need many tightly coupled threads/processes as in large MPI-based simulations or data analysis applications. Further clouds are clearly friendlier than current Grids and so provide an onramp or perhaps replacement for large scale Grid infrastructure such as the NSF TeraGrid. Clouds support smaller universities trying to enhance their Cyberinfrastructure related research and are natural places for educational laboratories. We suggest that the national academic infrastructure will evolve to a mix of clouds – possibly run by large campuses – linked to large MPI engines that may not change much from current supercomputer centers in the TeraGrid. Workflows will evolve to support the data intensive model of Hadoop and Dryad and access both classic clouds and MPI engines. We back this discussion with preliminary performance measurements from Eucalyptus and Nimbus.

Early results can be seen at <http://grids.ucs.indiana.edu/ptliupages/publications/CetraroWriteupJan09_v12.pdf> and <http://grids.ucs.indiana.edu/ptliupages/presentations/eScience_gcf_December08.pptx>