Production Distributed Infrastructure (PDI)

Daniel S. Katz (dsk@ci.uchicago.edu)

Senior Fellow, Computation Institute (University of Chicago & Argonne National Laboratory)

(all the following are personal opinions only)

State of the Art in PDI

- PDI - a distributed set of computational hardware and software that is intended to allow multiple people who are not the developers of the infrastructure to do something useful
- Three types of PDI exist:
  - Academic/Public Production (aimed at science)
    - TeraGrid/XSEDE, DEISA (HPC), OSG, EGI (HTC), Open Science Data Cloud (cloud), etc.
  - Academic/Public Research (aimed at computer science)
    - Grid5000, DAS, FutureGrid, PlanetLab, etc.
  - Commercial (aimed at whoever will pay)
    - Amazon (IaaS), Microsoft (Paas), perhaps SGI, Penguin (HPCaaS)
- In addition, lots of other components that may not be integrated together:
  - Campus clusters (HPC), campus Condor pools (HTC)
- All seem to do what they were designed to do reasonable well

- This view is mostly by funding and purpose
- Could also view by interface:
  - Command-line, grid (Globus, Condor), cloud

Open Challenges (in R&D, code, support, and/or policy)

- Goal – deliver maximum science (at minimum cost?)
  - Note – sustainability seems to be a hot topic, but it seems to be defined as: useful work should continue to be done, with someone else paying for it
- 2 views of this?
  - What are the components of infrastructure?
    - Hardware (nodes, interconnect, storage, network), software (system software, middleware, tools/libraries, applications), training (material, people), support (people), integration into PDIs
  - What's the vision of "the" PDI?
- Issues
  - how measure delivered science
    - We don't really know how to do this – we can measure papers (w/ a small time delay) or citations (w/ longer time delay)
  - How to develop the infrastructure/tools that will best do this
    - I think we are doing reasonably well at this at an individual infrastructure/tool level—we identify things that users want to do, then provide tools/systems that let them do these things — but it's not clear that we ever solve the whole integrated problem
  - How to integrate them
    - We can do this on a piece-by-piece basis, but don't really know how to do this in general (maybe because we don't have a single vision for where we are going)

Path Forward

- Greatest need is a single vision that defines the overall goal
  - NSF has CIF21 (http://www.nsf.gov/pubs/2010/nsf10015/nsf10015.pdf) which calls for such a vision, and 6 tasks forces that have created reports
  - DOE seems more focused on single large systems, and less on integrating them into an infrastructure
  - Some other agencies looking at clouds
- and is flexible enough to allow groups to work towards that goal in various ways
  - Should define metrics (to enable progress to be measured and see which work is most useful)
  - And an overall architecture with interfaces
    - OGF is trying to do some of this, but without sufficient US buy-in

Acknowledgements

- First slide (and probably some other thinking)based on chapter 3 of upcoming book:
  - S. Jha, D. S. Katz, M. Parashar, O. Rana, and J. Weissman. Abstractions for Distributed Applications and Systems. Wiley.
- And technical report that's coming sooner:
  - D. S. Katz, S. Jha, M. Parashar, O. Rana, and J. Weissman. Distributed Cyberinfastructures. Technical Report. Computation Institute, University of Chicago & Argonne National Laboratory, URL/number TBD
- And related eSI research themes (DPA and 3DPAS)