

# High Throughput Computing, Grid Computing, Cloud Computing, Etc. Definitions & Thoughts



Jay Boisseau  
Texas Advanced Computing Center  
July 16, 2008



THE UNIVERSITY OF TEXAS AT AUSTIN  
**Texas Advanced Computing Center**



## Some “Definitions”

- But before the definitions, some ***disclaimers***:
  - These my my own working definitions
  - I provide them to give context to our presentations, topics, activities, and projects
  - feel free to disagree with them!



2  
High Throughput Computing, Etc.



## Some “Definitions”

- Advanced Computing – a very generic term for *any* kind of computing beyond mainstream consumer /business computing. Generally implies something about scale:
  - Fastest components
  - Aggregating lots of componentsSometimes means ‘just’ advanced software, but usually implies advanced hardware
- Example: Texas Advanced Computing Center supports HPC, visualization, big data storage, high throughput computing, data analysis, etc.



3

High Throughput Computing, Etc.



## Some “Definitions”

- Cyberinfrastructure: the *integration* of potentially diverse computers, displays and visualizations, data, storage systems, instruments, sensors, etc. via software and networks (and policies/procedures?) to
  - Provide comprehensive capabilities
  - Provide aggregate capabilities (sometimes)
  - Share resources (sometimes)
- Example: NSF TeraGrid provides ‘high end’ cyberinfrastructure (spanning 12 institutions) for computational research in the USA



4

High Throughput Computing, Etc.



## Some “Definitions”

- Supercomputing, Parallel Computing, High Performance Computing (HPC):
  - Supercomputing: older term, originated when Cray and CDC supercomputers (1960-80s) were vastly more powerful than any other computers. Still used for any kind of tightly coupled, large-scale parallel computing systems
  - Parallel computing: aggregating computers or processors to address problems by dividing up the work
  - HPC: now, refers to parallel computing at large scale



5

High Throughput Computing, Etc.



## Some “Definitions”

- Distributed Computing: very general term, includes *any* kind of computing spanning more than one computer connected by a network. We all used distributed computer every day (connecting to the Internet for web, email, etc.).
- Examples: everything from email to web browsing to file sharing to B2B applications to clusters to remote visualization to grids to clouds to....



6

High Throughput Computing, Etc.



## Some “Definitions”

- Grid Computing
  - Grid Computing: using distributed computing with ‘standards’(?) to enable using multiple, often geographically distributed, computing resources, for such reasons as:
    - Resource sharing
    - Data Sharing
    - Workflow applications
    - Aggregating multiple systems (coupled applications, “metacomputing,” high throughput computing, etc.)
  - Examples: usually done for ‘science/research’ but also for business, with examples like DEISA (Euro data grid), World Community Grid, etc.



7  
High Throughput Computing, Etc.



## Some “Definitions”

- Cloud Computing: currently the most hyped term, refers to using remote, generally massively aggregated, computing resources transparently for local ‘work’/purposes with high throughput/performance. Enabled by ‘standards,’ virtualization (sometimes), and massive data centers (e.g. Google, MS, etc.)
- Examples: Amazon’s EC2 compute service; Amazon S3 storage service; Apple’s MobileMe (storage, file sharing, etc.); Flickr, YouTube; etc.



8  
High Throughput Computing, Etc.



## Some “Definitions”

- High Throughput Computing: using distributed computing (potentially grid computing) to enable lots of jobs to be scheduled to available resources to complete as fast as possible. Term was popularized by Condor project (Miron Livny, University of Wisconsin). Arguably refers to ‘volunteer grids’ as well
- Examples: Condor flocks, World Community Grid, LHC project, Open Science Grid, etc.



9

High Throughput Computing, Etc.



## High Throughput Computing

- Goal: to integrate multiple computing systems to enable large numbers of tasks to be scheduled and completed as rapidly as possible
- Example technologies: Condor, BOINC, SGE
- Resources: can be centrally managed servers (clusters, clouds) and/or distributed PCs
  - Amazon EC2 provides cycles, using user machine images (Xen virtualization), via web services
  - World Community Grid leverages 1M+ volunteer PCs (including even laptops)



10

High Throughput Computing, Etc.



## Value of High Throughput Computing

- Majority of computational science is performed on workstations/PCs
  - HPC systems not needed for all tasks
  - Many tasks need to be repeated—a lot
    - Running simulations with to explore a parameter space
    - Running simulations on different data sets
    - Analyzing experimental results that ongoing/repeated
    - Etc.
  - HTC tools enable this, sometimes from the desktop (e.g. Condor)



## “Many-Task Computing”

- Many-task computing is a new term, arguably the same as high throughput computing: “the execution of independent, sequential jobs that can be individually scheduled on many different computing resources across multiple administrative boundaries” (Ian Foster)
- Primary difference (?) is the scale of the tasks is much shorter, increasing emphasis on scalability of the enabling infrastructure



## Some References

- Condor Project
  - <http://www.cs.wisc.edu/condor/>
- Many Task Computing Workshop, Nov 2008:
  - <http://dsl.cs.uchicago.edu/MTAGS08/>
- World Community Grid
  - <http://www.worldcommunitygrid.org>
  - We encourage you to download the client, start a team!
- BOINC Project:
  - <http://boinc.berkeley.edu/>



13  
High Throughput Computing, Etc.



## Some Cloud Projects

- Eucalyptus
  - <http://eucalyptus.cs.ucsb.edu/>
- Virtual Workspaces
  - <http://workspace.globus.org/>
- Grid Appliance
  - <http://www.grid-appliance.org/>
- Open Nebula
  - <http://www.opennebula.org/doku.php>
- Amazon Elastic Cloud (costs \$, but very little)
  - [aws.amazon.com/ec2](http://aws.amazon.com/ec2)



14  
High Throughput Computing, Etc.

