# Lecture 21:
# Grids and Clouds

David Bindel

11 Nov 2011

# Logistics

- Project 3 due Monday at midnight
  - I will be traveling Sunday – ask questions soon!
- Final project:

  12/1: Short presentation

  12/16: Final reports
- Today: Joint presentation with Tao Zao

# Project 3 comments

- Second MPI implementation should be memory scalable!
  - May want to think about how to do initialization...
- 1D ring doesn't save on communication volume
  - 2D layout would be better – see dense LA lecture
  - But this exercises what I want you to learn!
  - And you overlap communication with computation
- Be careful to communicate about termination!
  - `MPI_Allreduce` works...

# Grids

http://en.wikipedia.org/wiki/File:
Electric_transmission_lines.jpg

# Clouds

http://en.wikipedia.org/wiki/File:Cloud_in_nepal.jpg

# Portals


http://en.wikipedia.org/wiki/File:
Portal_standalonebox.jpg

# Watch out, little guy!


# Utility computing

Names change, but the concept is attractive:

- Flexible access to compute time and data storage
- Maybe not in a single administrative domain
- Using simple, standardized interface
- Maybe with nice high-level interfaces

# Cycle scavenging

- Condor project (1988-present)
  - Idea: Use idle cycles on networked computers
  - Support for transparent checkpointing and migration
  - Now managing EC2 Spot Instances!
- Volunteer computing
  - SETI@Home (1999-present)
  - Folding@Home (2000-present)
  - BOINC (2003-present)
- Good for high throughput in embarrassingly parallel settings
- Not so good for solving PDEs...

# Globus (1996-present)

Dream: uniform access to distributed

- Compute power
- Data storage
- Data sources (satellites, instruments, etc)

Used by Teragrid / XSEDE.

Some components:

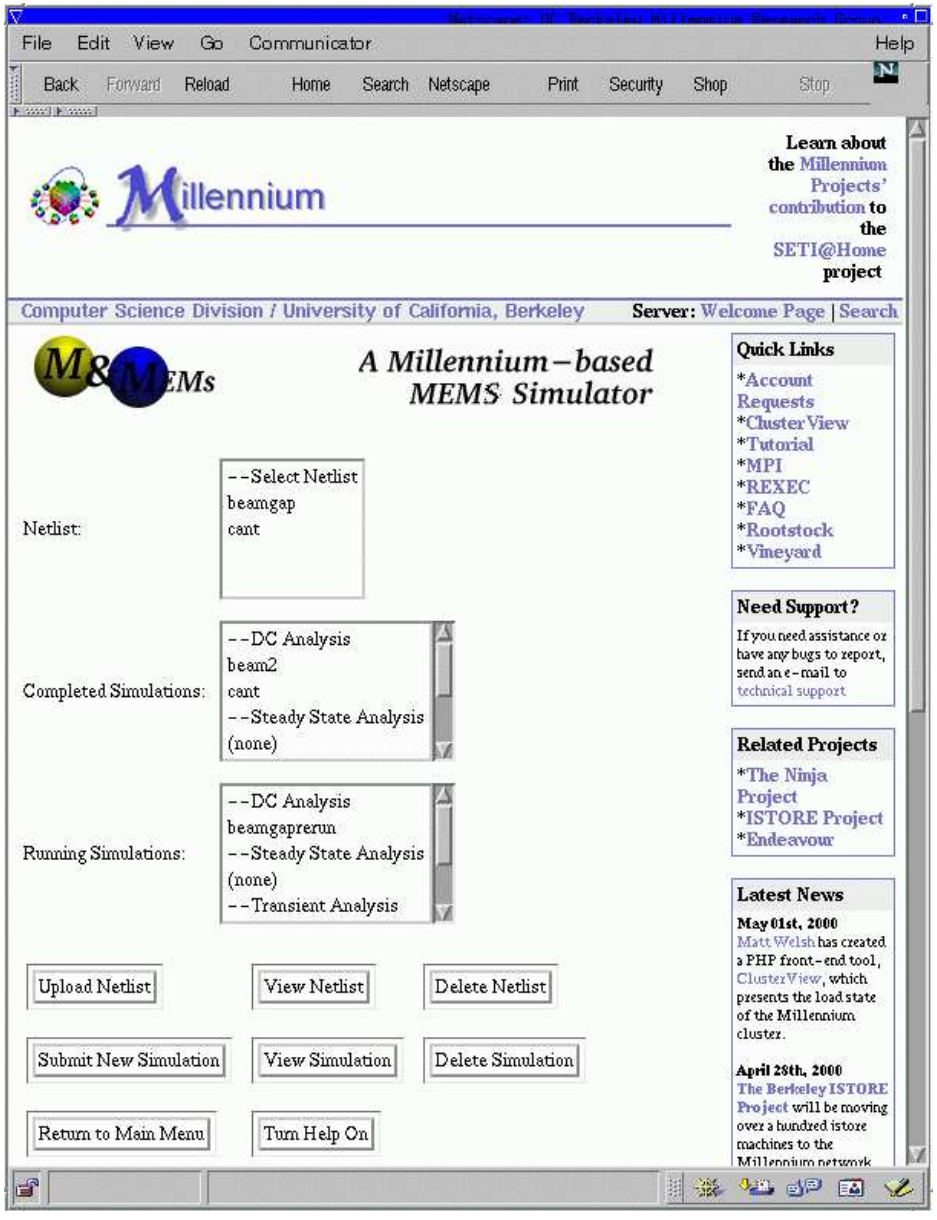
- Grid Security Interface (GSI)
- Grid Resource Allocation and Management (GRAM)
- MPIg (aka MPICH-G4)

# Gateways / portals

Remote access interfaces (often via web) to science-specific tools:

- XSEDE (NSF) lists several
- hpc2 (NYSTAR – NY state) hosts several
- Nanohub hosts several
- NERSC hosts several
- ...

# M&MEMS: A personal recollection (2000)



# Cloudy prospects

Why not run lots of HPC on EC2?

- Have to start worrying about individual node failures
  - Will be a worry for anyone if we succeed at exascale...
- Communication costs are a killer

Partial solutions:

- Better algorithms (communication avoiding)
- New programming frameworks?

# “Mid-range” computing on clouds

http://www.nersc.gov/assets/StaffPresentations/2011/MoabCon-Canon-Cloud-presented.pdf

Paper: www.lbl.gov/cs/CSnews/cloudcomBP.pdf