Workshop Day-1 Summary

**Large Scale Computing and Storage Requirements
for Biological and Environmental Research
Joint BER / ASCR / NERSC Workshop**

NERSC

Lawrence Berkeley National Laboratory

May 7-8, 2009

Summary

- Users need for more resources for DOE SC computing
  - But we need more concrete, science-based justification
- Need for predictable throughput
  - Microbial Genomics and GFDL ESM
  - Need to differentiate between real-time needs and higher desired batch turnaround
    - Slow batch turnaround time may be because of queue policy or because of insufficient resources

Summary

- Data management issues transcend all science areas, multiple projects
  - Exponentially increasing
  - Some disagreement on volume at a single site

Summary

- Users see need for guidance / help in programming models for multicore
  - Need training, advice, consulting support, deployment of software choices on production machine

Summary

- Better support for workflows
  - Kepler/SDM
  - Microbial informatics special workflow needs
  - Computational "beamline" workflow & reliable turnaround, production-line rate
  - ESG

Summary

- Portals and data hosting: dynameomics.org as an example
  - ESG, GFDL, subsurface effort soon
- I/O rates need improvement
  - Need to drill down further
  - Need ASCR investment to maintain infrastructure
- Use coupled simulation as a benchmark

Summary

- Expanded view of pentagon
- Clear support for staggered deployment of NERSC systems

Summary

- Improved system stability
  - What can NERSC do improve system
  - What can NERSC do to help users mitigate
    - Community needs more training in fault tolerance
- Low-latency MPI messages
- Problems scaling CG
  - ALL_Reduce performance
  - Wonder about other solver choices

Some Recommendations

- What are the specific scientific problems that can be solved with significant improvements in computing?

- Need to be sure to link specific computational needs to scientific discoveries within 5-year timeframe.
  - "Increased resolution" too vague