Notes

• Group project proposal questions?
  • proposals due Monday

• Swati working on grading OpenMP project
Computational Grids

• Infrastructure for dependable, consistent, pervasive access to high-end compute resources

• Why?
  • Increased compute power to users
  • Demand-driven access to compute power
  • Utilize idle capacity (desktop grids)
  • Sharing of computational results (data grids)

• Application areas
  • Distributed supercomputing – to aggregate compute power
  • High throughput – desktop grids, volunteer computing
  • On-demand computing – remote servers, instruments, etc.
  • Data intensive computing – LHC, SDSS, remote sensing

• Grid users
  • Infrastructure developers
  • Tool developers
  • Application developers – various programming models
  • End users of grid-enabled applications
  • System administrators – across admin domains
Grids (cont.)

• **Grid architecture**
  - End systems – individual components
  - Clusters – mostly homogeneous, larger scale
  - Intranet – distributed, heterogeneous, limited centralized control
  - Internet – heterogeneous, no centralized control, security and trust issues, more geographic distribution

• **Current status is that grids are now everywhere, for high performance computing applications, for large-scale volunteer computing projects**
  - but the infrastructure is still somewhat brittle, and has not had as widespread impact on science and other areas as was expected 15-20 years ago
  - A lot of it has morphed into cloud computing (one big difference is virtualization)
Data Grids

• Key issue is accessing, managing, and computing on large, distributed data collections
  • geographic distribution
  • high performance demands

• Key design principles
  • Mechanism neutrality – independence from how data is stored, metadata is stored, data transferred, etc.
  • Policy neutrality – expose performance-related decisions to users if they want to make them
  • Grid infrastructure compatibility – use existing Grid infrastructure (they mean Globus toolkit) for basic services
    • authentication, resource management, resource discovery
  • Information infrastructure uniformity – easy access to resource structure and state
    • to enable applications to adapt to current system conditions
Data Grids (cont.)

• Core services
  • Storage system abstraction – create, destroy, read, write, manipulate file instances
    • could be data in a DBMS, or a mass-storage system, or data accessed via an http server
  • Data access – including 3rd party transfers
  • Metadata service – to publish and access info about file instances in a uniform way – stored in a metadata catalog
    • structured as a hierarchical, distributed system (LDAP)
  • Authentication/authorization via Grid Security Infrastructure (GSI)
  • Resource reservation and co-allocation for storage systems and networks
  • Performance measurement and instrumentation

• Replica management as a higher level service
  • to create and manage file instance copies
  • problems addressed include replica selection and data filtering