

Earth System Grid II

Turning Climate Datasets into Community Resources

Participating Institutions: **ANL**, USC,
LLNL, **NCAR**, NERSC, ORNL

ORNL Staff: **David Bernholdt**, Kasidit
Chanchio, Trey White

The Problem

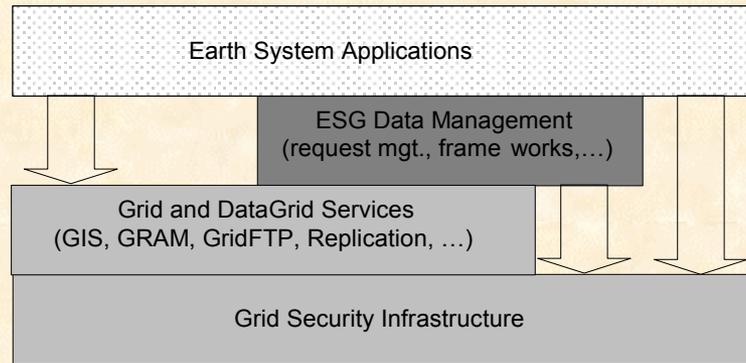
- Community access to, and analysis of, climate data
- Climate unique among Grand Challenge problems in terms of length & number of simulations, and volume of model output requiring subsequent analysis
 - One century simulation requires a month and produces ~10 TB archival output
- Data archived and analyzed by PCMDI
- Extracts and derivatives of data also used for different purposes by numerous researchers around the nation

The Earth System Grid

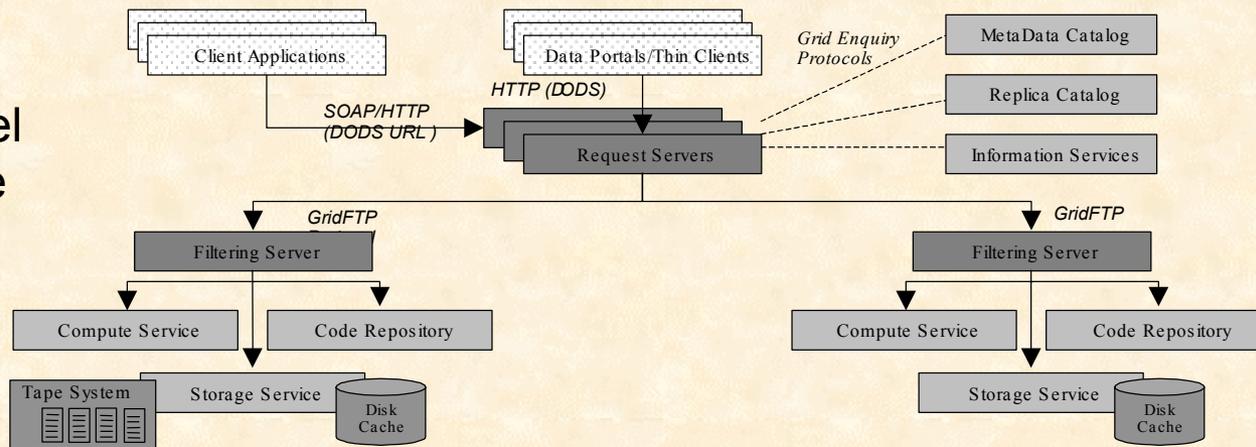
- Integrated infrastructure for data storage management, movement, and reduction
- High-speed data movement built on top of GridFTP
- High-level replica management & caching to improve throughput and response time
- Security model to allow access while enforcing policy constraints
- Remote data access via Community Data Portal and DODS
- Intelligent request management
- Filtering servers to permit processing and reduction of data closer to point of residence

ESG-II Architecture

High-Level Design



Service Level Architecture



Grid Infrastructure Services

- Built on top of core Grid services
 - Globus Security Infrastructure (GSI)
 - Metacomputing Directory Service (MDS)
 - Globus Resource Allocation Manager (GRAM)
- Catalog Services
 - Metadata catalogs, map attribute names/values to specific data files
 - Replicas catalogs maps data file names to physical storage locations
 - Software catalogs provide access to related executables
- Data Management Services
 - Data transport built on top of GridFTP
 - Data replication services performs transfers, insures consistency

User/Application Interaction

- Filtering Servers
 - Currently can only request entire files, file subsets, results from simple pre-installed filters. ESG-II will be more flexible
 - Analysis functions may be pre-defined or user-supplied, composable
- Request Management Services (RMS)
 - Mediate between user request and sequence of interactions on Grid necessary to satisfy
- Data Portals
 - Web-based or thin clients, support w/ DODS RMS
- Heavy-Weight Applications
 - Provide APIs for data staging, caching in distributed env.

ORNL's Role

- ORNL will be an early deployment site for ESG-II
- Center for Computational Sciences is a leading location for climate modeling runs
- Probe Storage Testbed (joint w/ NERSC) provides unique facility for climate data storage and movement

