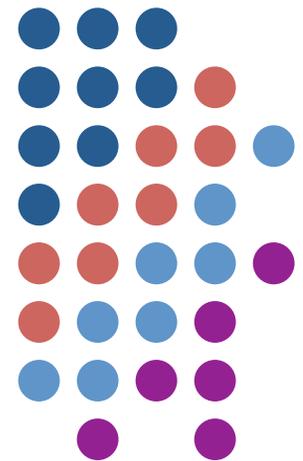


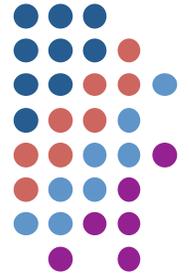
GLADE Overview

GLobally Accessible Data Environment

CHAP Meeting
Pamela Gillman
29 April 2010



Outline



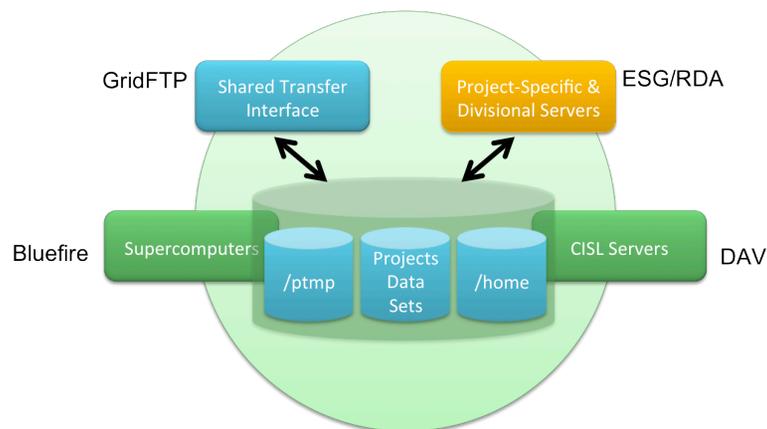
- Primary Objective
- Architecture and Benefits
- Project Timeline
- Space Allocations
- Allocation Policies
- Futures



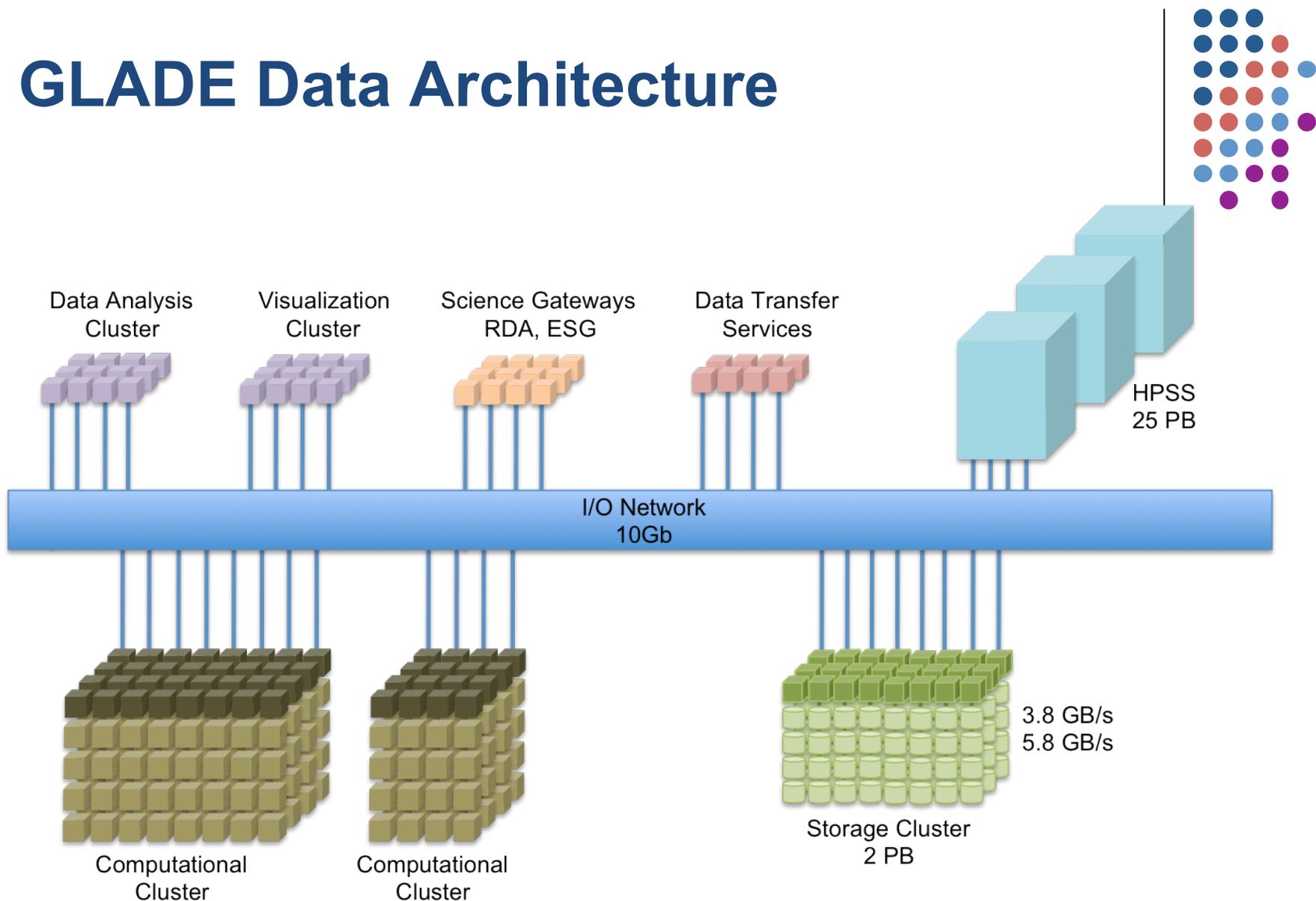
GLADE Primary Objectives



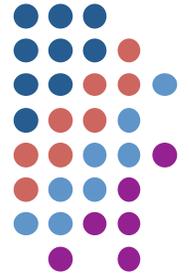
- Unified and consistent data environment for NCAR HPC
 - Supercomputers, DAV, and storage
 - Shared transfer interface and support for projects
 - Support for analysis of IPCC AR5 data
 - Service Gateways for ESG & RDA data sets
- Data is available at high bandwidth to any server or supercomputer within the GLADE environment
- Resources outside the environment can manipulate data using common interfaces
- Choice of interfaces supports current projects; platform is flexible to support future projects



GLADE Data Architecture



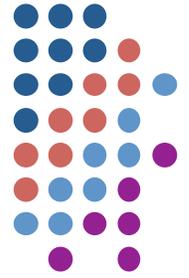
GLADE Benefits



- Information centric
 - Data can stay in place through entire workflow
 - Access from supercomputing, data post-processing, analysis and visualization resources
 - Direct access to NCAR data collections
- Availability of persistent longer-term storage
 - Allows completion of entire workflow prior to final storage of results either at NCAR or offsite
- Provides high-bandwidth data transfer services between NCAR and member institutions

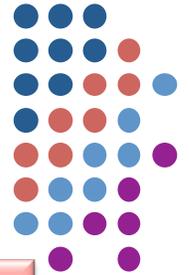


Project Timeline



- Phase I – [April 30]
 - Installation of high-bandwidth infrastructure
 - Installation of storage infrastructure
 - Installation of initial data transfer gateways
 - Configuration of ESG & RDA science gateways
- Phase II – [July 30]
 - Expansion of storage capacity (total capacity of 1.6 PB usable)
 - Expansion of data transfer services
 - Complete security implementation on DAV servers and bluefire
 - Add access to GLADE from frost/twister and lynx
- Phase III – [September 30]
 - Augmentation of connectivity to bluefire compute nodes
 - Expansion of IB infrastructure to include visualization cluster
 - Expansion of data transfer services if necessary

GLADE Data Environment



Project based space allocations

- Allocated through project space request process
- We do not scrub, no automated backups

Scratch space

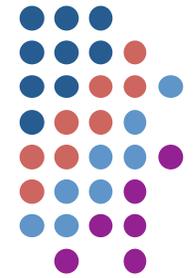
- User based quotas
- Space is scrubbed, no exceptions, no automated backups



Data Collection spaces

- No automated backups

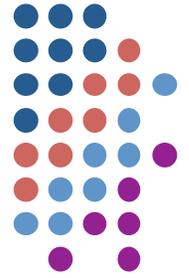
Total Storage Allocation Overview



Space	Available	Allocated	Free	Description
data01	512 TB	512 TB	0 TB	ESG hosted data collections, AR5/CMIP5 work space
data02	496 TB	496 TB	0 TB	RDA hosted data collections
project	680 TB	302 TB	378 TB	Project space
scratch	224 TB			Additional scratch work space



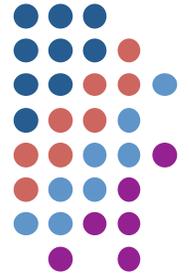
Current Allocation Request Process



- scratch quota increase
 - User may request increase up to 500GB
 - Needs greater than this need to request a project space
- Project Space Allocation
 - User Submits Request Form
 - CISL reviews the request
 - < 10TB – DASG may approve
 - 10TB-20TB – OSD director may approve
 - > 20TB – CISL Council may approve



Future Allocation Request Process

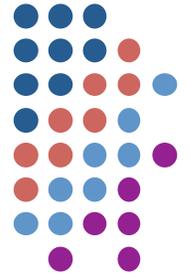


- Committee to evaluate larger requests
- Well defined evaluation criteria based on scientific merit
 - Is having a compute allocation sufficient?
- Funding opportunities for larger requests
- Backup strategies for critical allocations
- Evaluation criteria for hosted data collections

- Other ideas / recommendations?

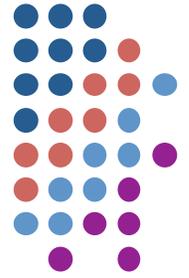


GLADE Futures



- Additional WAN mounts to select NCAR divisions
 - CGD, MMM
- Addition of storage to host a centralized home file system
 - Targeting 60TB
 - Will require new backup strategies
- Develop usage metrics





QUESTIONS?

