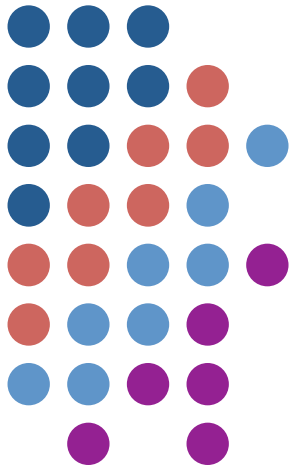
# NCAR's Globally Accessible Data Environment (GLADE)



Globus World 2014

16 April 2014

Pamela Gillman, NCAR

Manager, Data Analysis Services Group

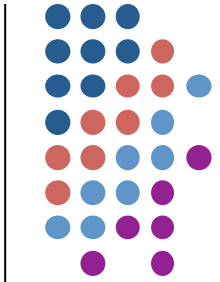
# Data Analysis Services Group

## NCAR / CISL / HSS / DASG

- Data Transfer and Storage Services
  - Pamela Gillman
  - Joey Mendoza
  - Craig Ruff
- High-Performance File Systems
- Data Transfer Protocols
- Visualization Services
  - John Clyne
  - Alan Norton
  - Scott Pearse
  - Miles Rufat-Latre (student)
- VAPOR development and support
- 3D visualization

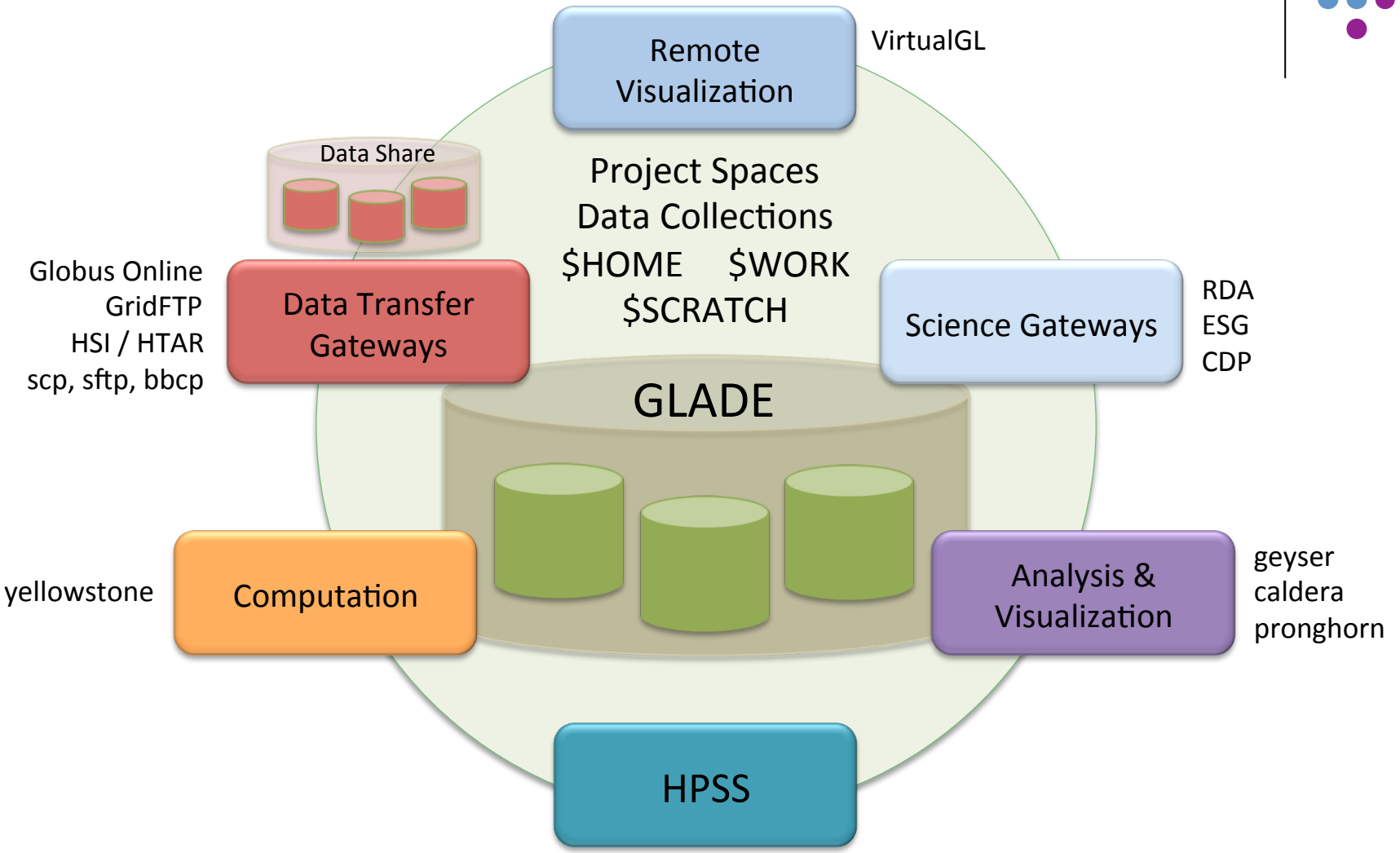
# GLADE

## GLobally Accessible Data Environment

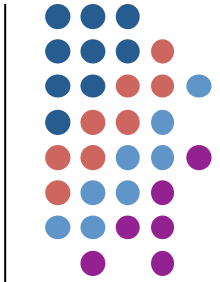


- Unified and consistent data environment for NCAR HPC
  - Supercomputers, Data Analysis and Visualization Clusters
  - Support for project work spaces
  - Support for shared data transfer interfaces
  - Support for Science Gateways and access to ESG & RDA data sets
- Data is available at high bandwidth 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 Environment

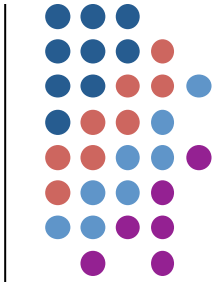


# GLADE Storage Overview



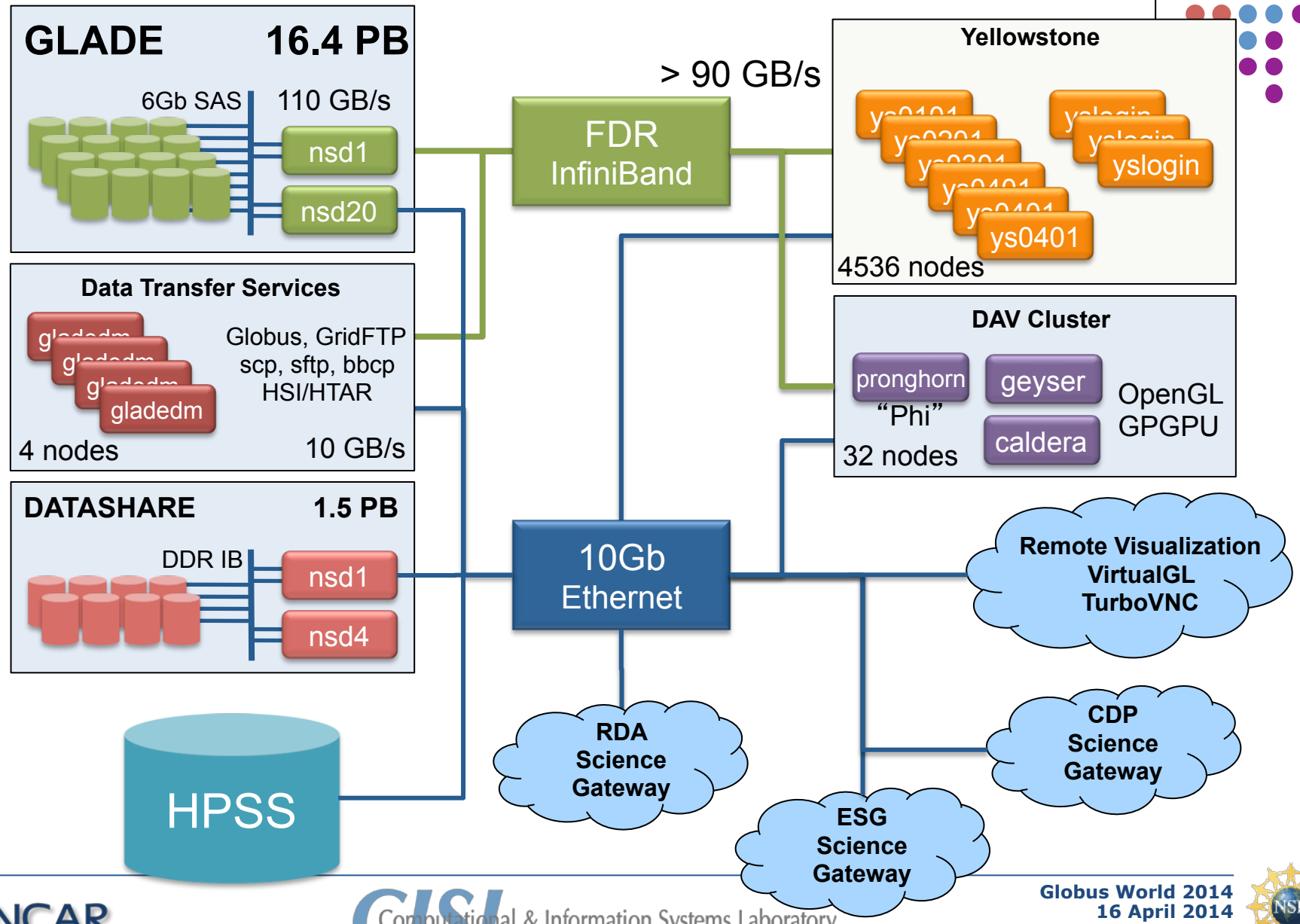
- 10.5 PB useable
  - + 6 PB useable, total 16.4 PB usable (April 2014)
- > 90 GB/s sustained bandwidth
- 76 DCS3700 systems with 1 expansion chassis
- 6840 3TB drives
- SAS direct connect to NSD servers
- 20 NSD servers, 6 management nodes
- 2 InfiniBand management nodes
- 4 data transfer nodes
- 1 108-port IB FDR 14 switch, 6 ethernet switches
- 21 racks
- GPFS Parallel File System

# DataShare Storage Overview



- 1.5 PB useable
- > 5 GB/s sustained bandwidth
- 1 DDN 9900 system
- 300 1TB drives, 900 2TB drives
- 4 NSD and management nodes
- DDR IB direct connect from storage to servers
- 3 racks
- GPFS Parallel File System
- Globus Plus integration

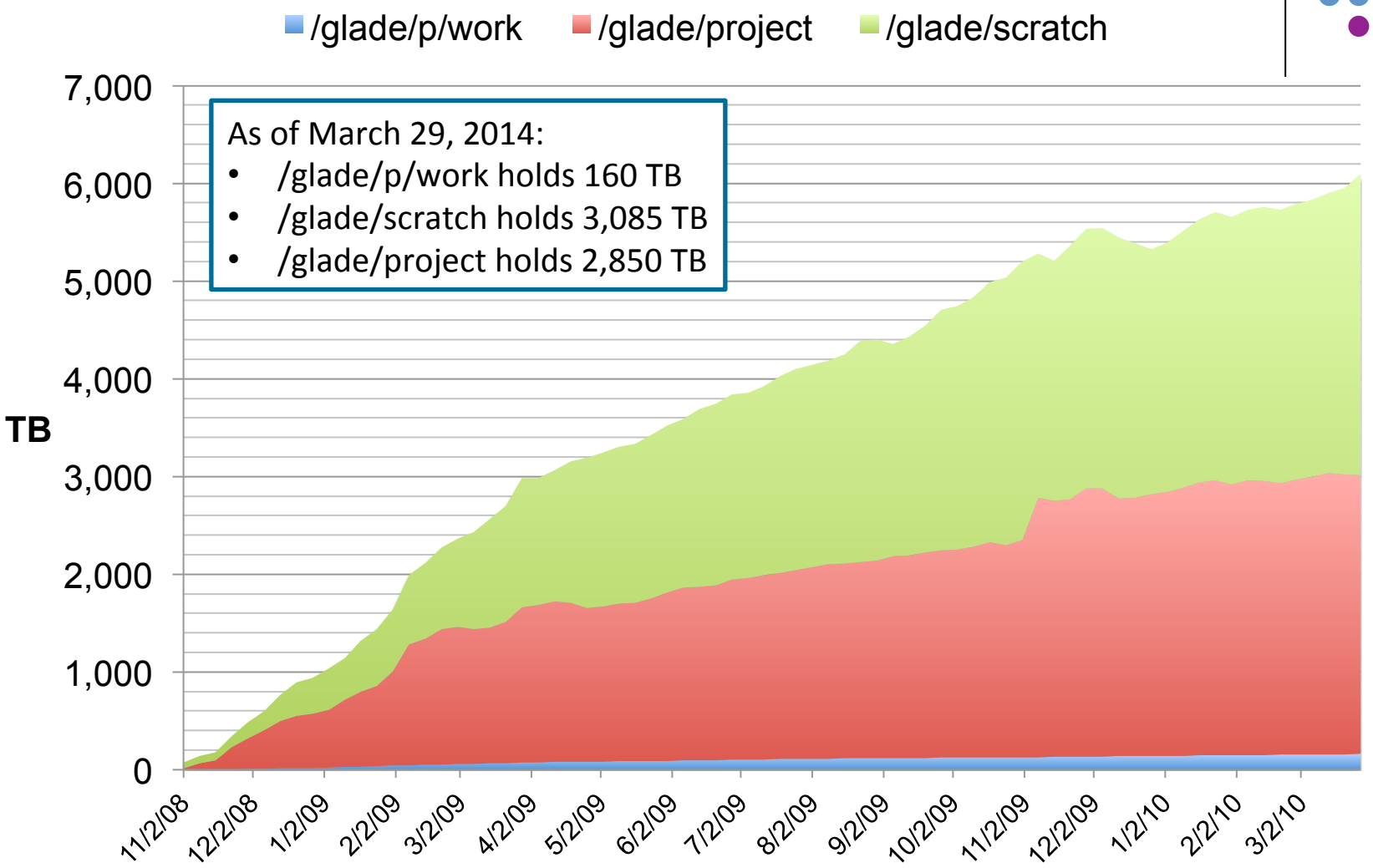
# Data Service I/O Networks



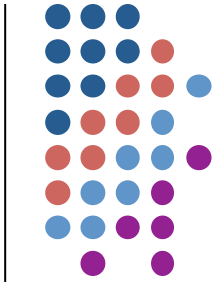
# GLADE Data Workflow Solutions

- Information centric data model
  - 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 and data sharing services between NCAR and peer institutions

# GLADE Growth

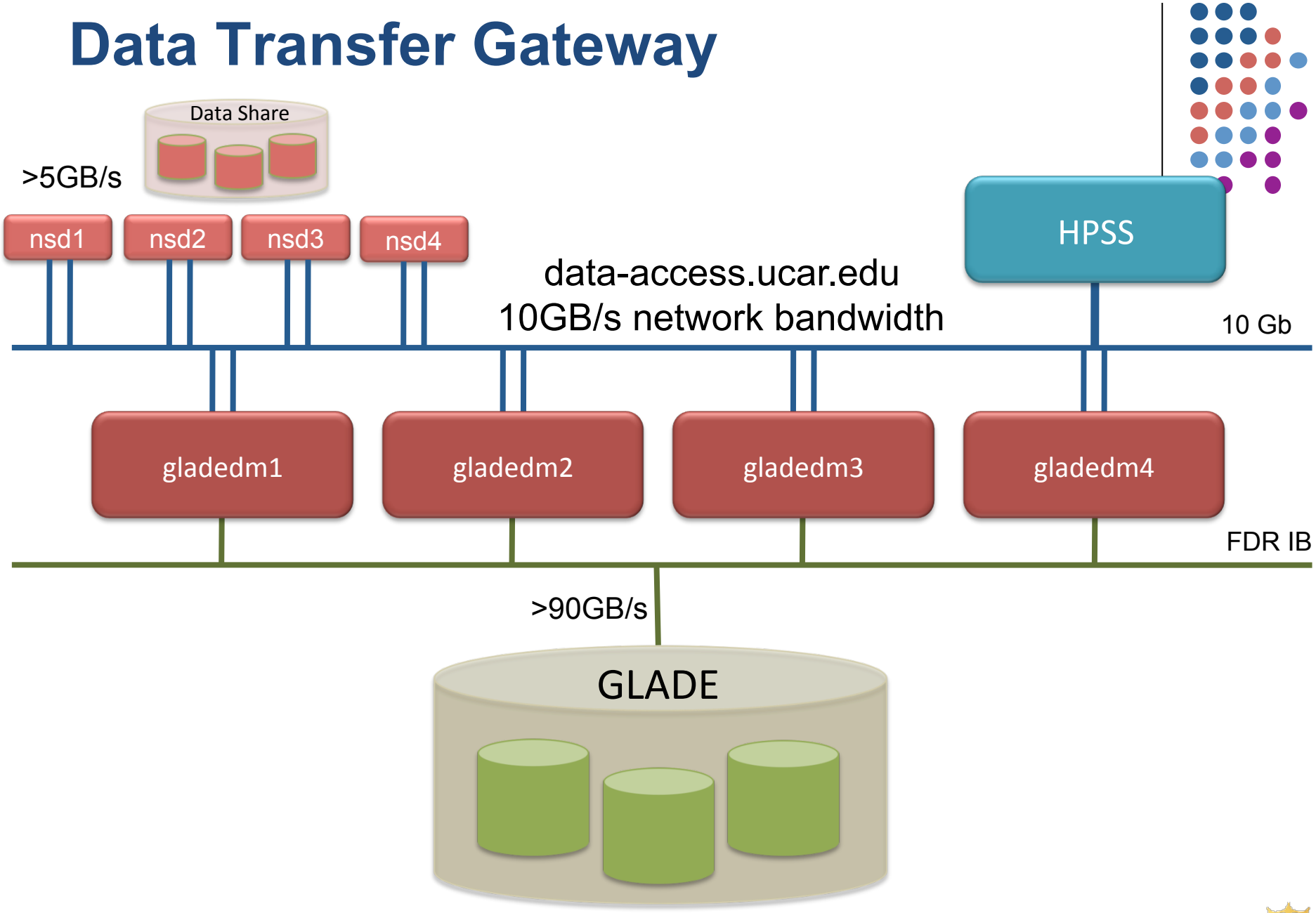


# CISL Science Gateway Support

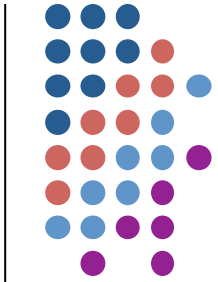


- Research Data Archive
  - 1 PB allocation
  - sub setting services are performed on geyser
  - direct access to online data collections from batch jobs
- Earth Systems Grid / Community Data Portal
  - 1 PB allocation
  - direct access to CMIP5 and NARCAP data from batch jobs

# Data Transfer Gateway

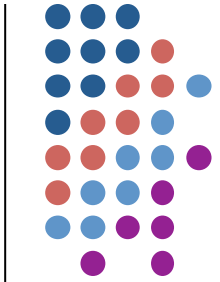


# Data Transfer Services



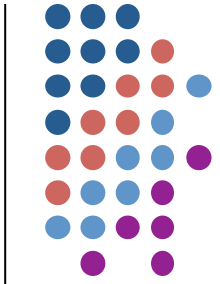
- Globus Online Endpoints
  - launch and forget data transfers
  - Access with users UCAS account and token
    - ncar#gridftp
  - Access with users XSEDE Account
    - xsede#ncar
    - xsede#glade
  - Web UI, CLI, REST API
  - Globus Connect for transfer to/from your desktop
- gridftp, globus-url copy, scp/sft, bbcp
- HSI/HTAR for HPSS access through LSF

# Data Sharing Services



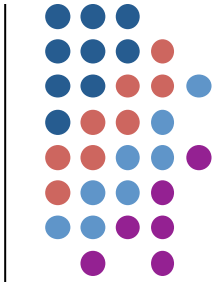
- *ncar#datashare*
- Globus Plus implementation
- data sharing allocations for self-publishing or data delivery
- data owner controls access
  - can create groups for access control
  - can share ‘read-only’ or ‘read-write’
- user can create custom access interfaces
  - CLI or REST API

# Data Sharing Use Cases

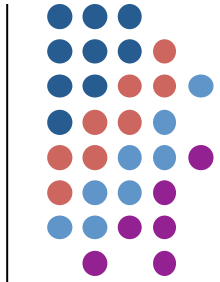


- Delivery of data from non NCAR users for publication in a Science Gateway
- Delivery of 3D visualization to non NCAR users
- Publication of supporting data associated with publication
- Share a file or data set with a non NCAR collaborator

# Data Sharing Futures



- User Outreach and Education
  - need more publicity for available services
  - schedule a user training seminar
- Potential to expanding ‘Sharing’ capability to the larger project spaces
- Potential to couple GlobusOnline more tightly with the Science Gateways
- Potential project to help build a custom UI for access to a data collection
- Re-evaluation HPSS integration with Globus



pjg@ucar.edu

# QUESTIONS?