Basic Idea

- Create a very large, scalable, parallel storage pool composed of generic storage blocks with software.
- With selective longer-term backup to another storage system.
- Using multiple IO nodes that communicate to the pool via parallel links to supply:
  - CIFS/SMB file service
  - Single-channel NFS file service
  - web access for distributing campus data resources
  - direct parallel access to clients requiring very fast IO
  - high-speed data transfer via scp, sftp, rsync, Globus/GridFTP/DTN thru 10/40/100G links to CENIC

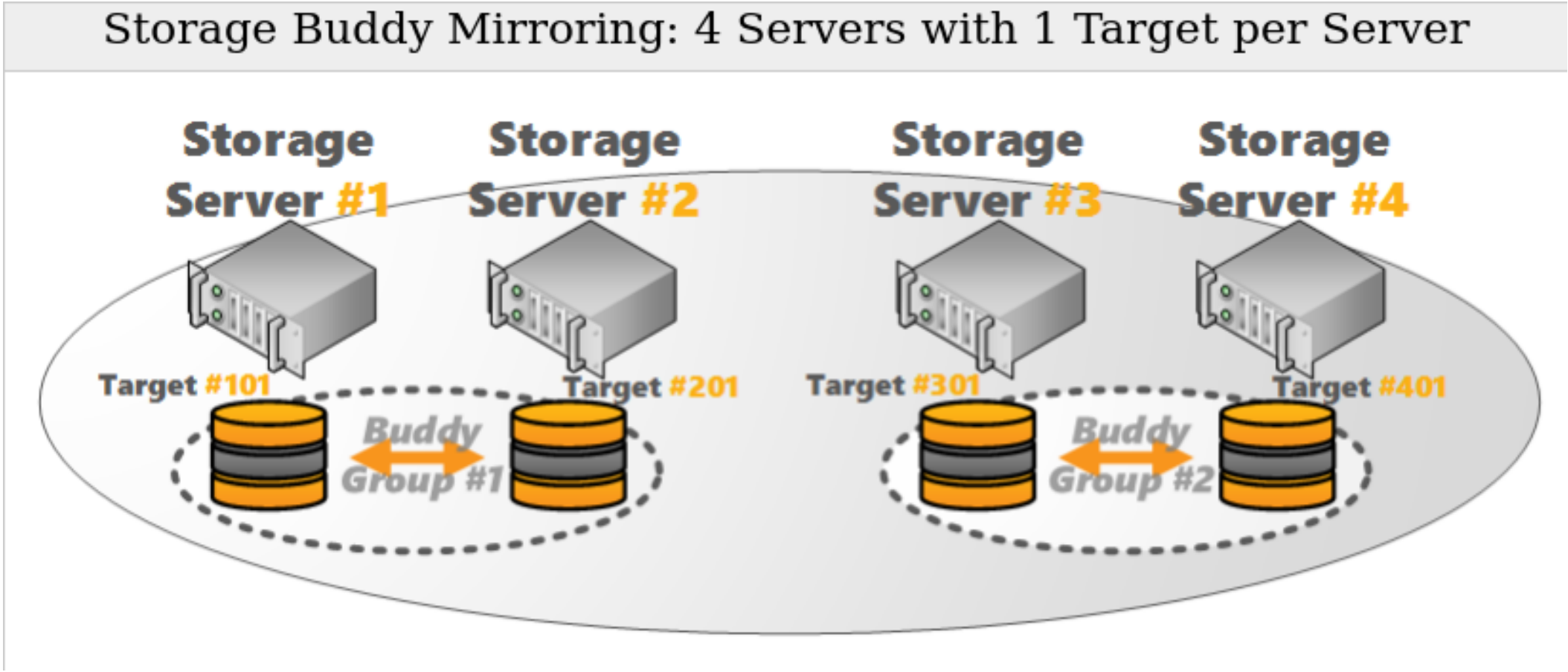
Approaches

- ~~Complete Replication~~
  - Copy-On-Write (COW) of entire filesystem
  - Dumb, wastes space & $ but saves time
- Mirror Specific Directories**
  - COW to specific dir trees
  - Omit large dumb dirs
  - More trouble than complete replication but saves space.
- ~~Selective Backup~~
  - Could add as an optional extra
  - More overhead, separate filesystems

Outline

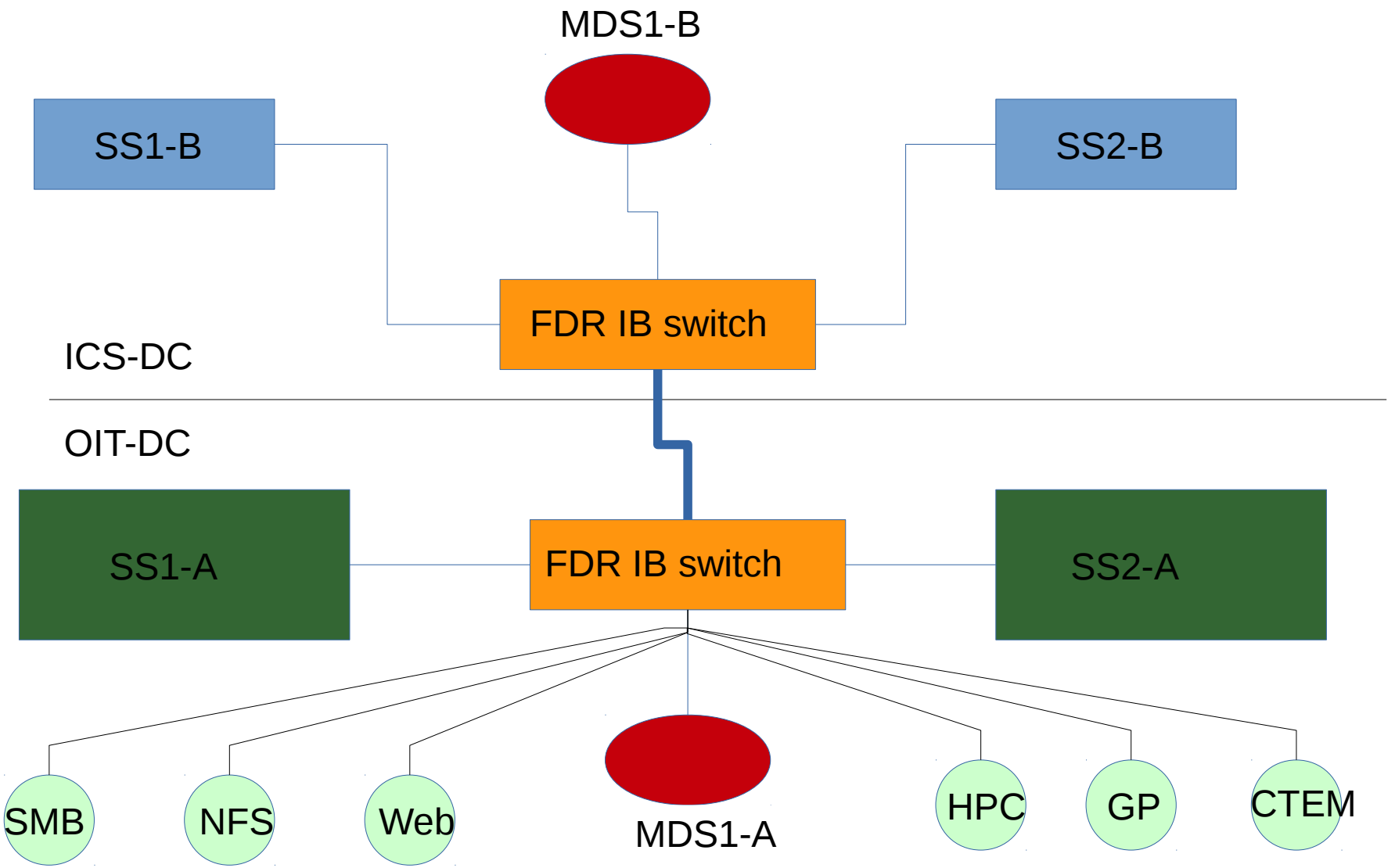
- Mirrored MD servers writing to NVME SSDs
- Mirrored Storage targets with Buddy group failovers.
- Primary servers are in OIT-DC
- The Buddy servers are in ICS-DC
- All comms over FDR IB
- Need dedicated switches. (Can’t piggyback HPC switches.)

BeeGFS Directory Mirroring



http://www.beegfs.com/wiki/AboutMirroring2015
http://www.beegfs.com/wiki/BuddyGroups2015

BeeGFS Buddy System



Multipath to JBODs