Science & Technology
Facilities Council

James Adams

Scientific Computing Department
Science & Technology Facilities Council
Rutherford Appleton Laboratory

Sirius and Echo

Two “cloudy” storage services run by SCD at RAL, both backed by Ceph object stores.

This talk will assume you know what Ceph is...

Different problems to solve

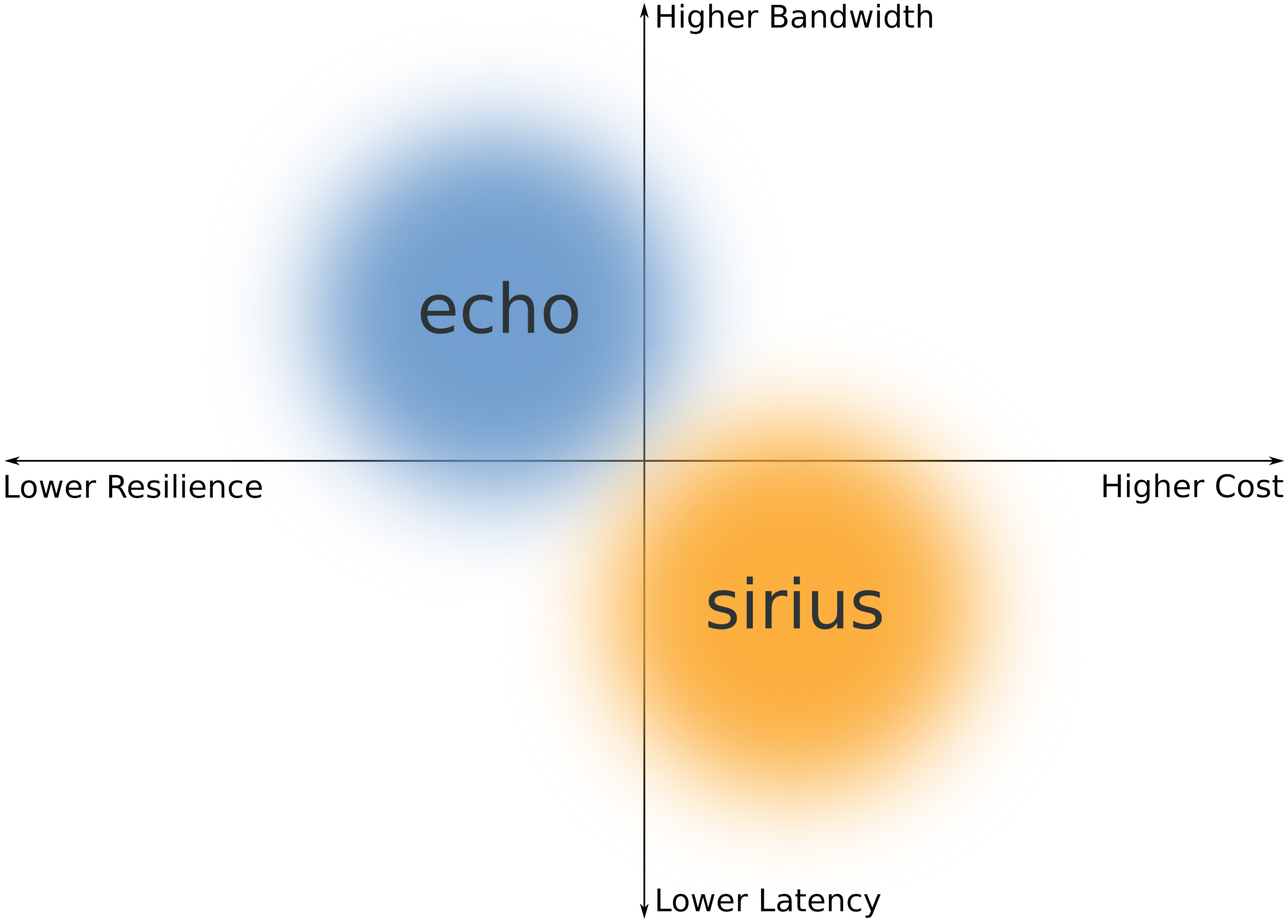
- Sirius
  - Provide low-latency storage to STFC private cloud
  - Host golden and running disk images for virtual hosts
  - Provide persistent block storage to services

- Echo
  - Disk only storage service to replace Castor for LHC VOs
  - Scale to meet data demands of LHC to 2020 and beyond
  - Provide industry standard access protocols

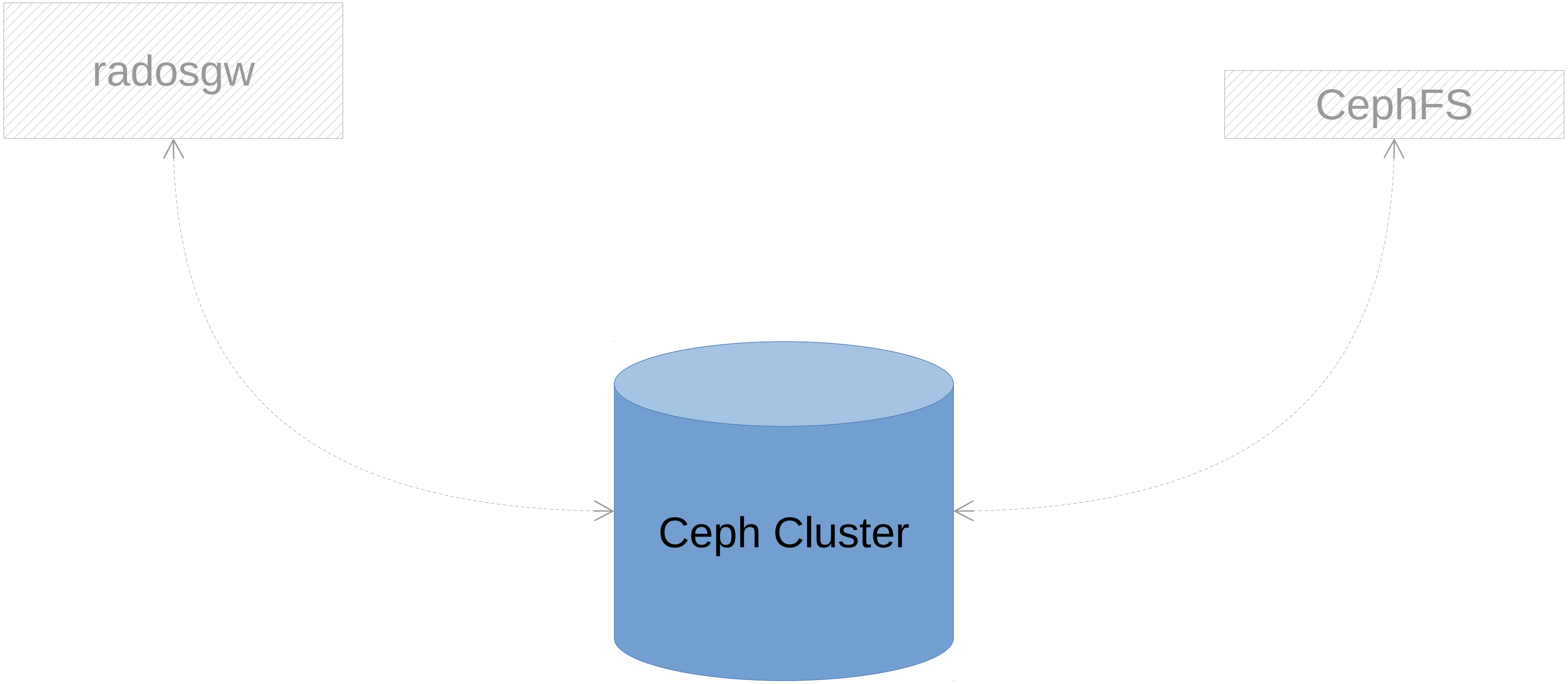
Different solutions

- Sirius
  - Block devices
  - Optimised for latency
    - Slim storage nodes
  - 750TiB raw storage
    - More being purchased
  - Replicated pools (3x)
    - Cost similar to service nodes
  - Individual users
  - Internally facing
    - STFC funded

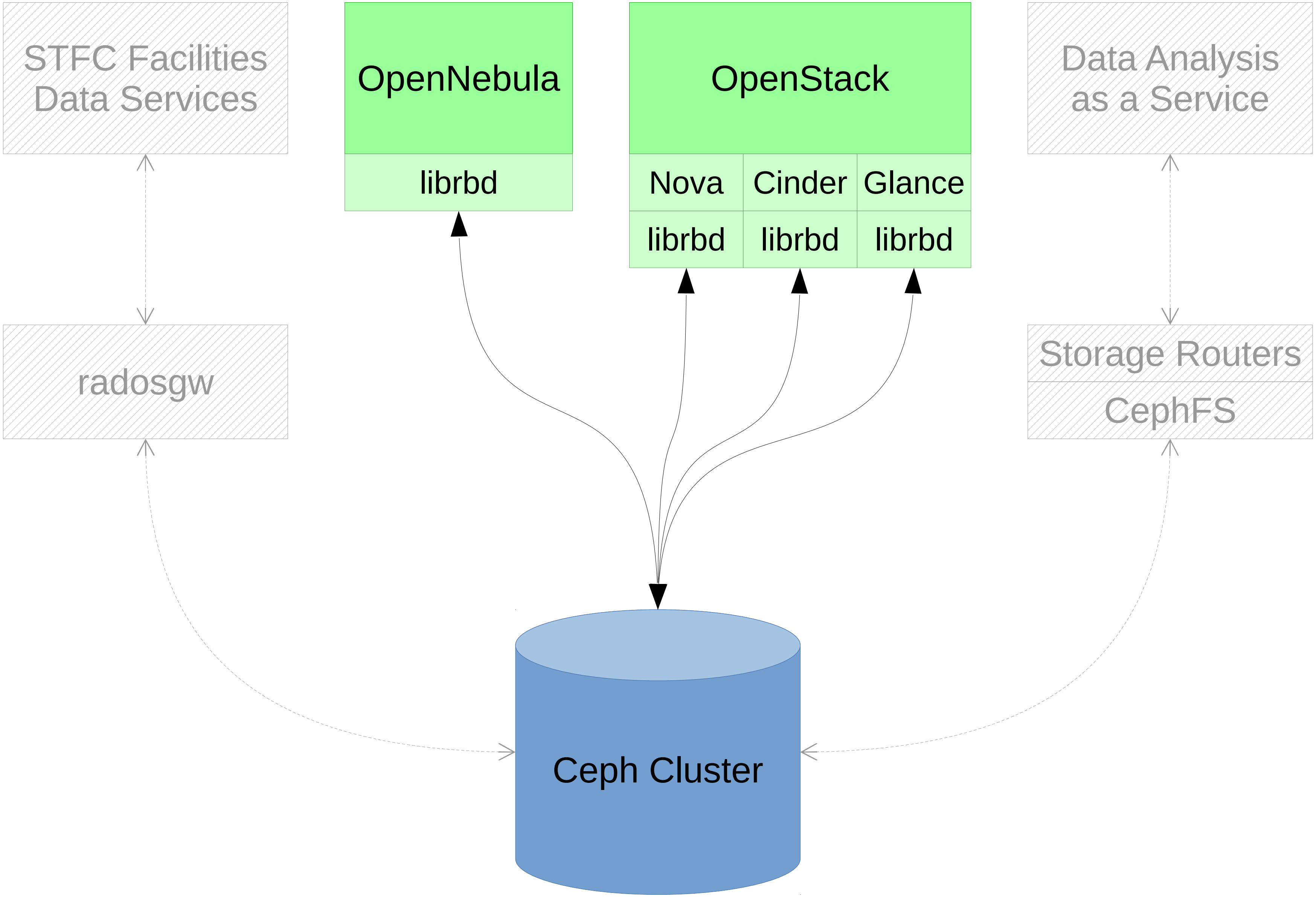
- Echo
  - File/object storage
  - Optimised for bandwidth
    - Fat storage nodes
  - 4PiB raw storage
    - Additional ~13PiB procured
  - Erasure coded pools
    - Cost similar to CASTOR
  - VOs/Science communities
  - Externally facing
    - GridPP funded


Higher Bandwidth
echo
Lower Resilience
Higher Cost
sirius
Lower Latency

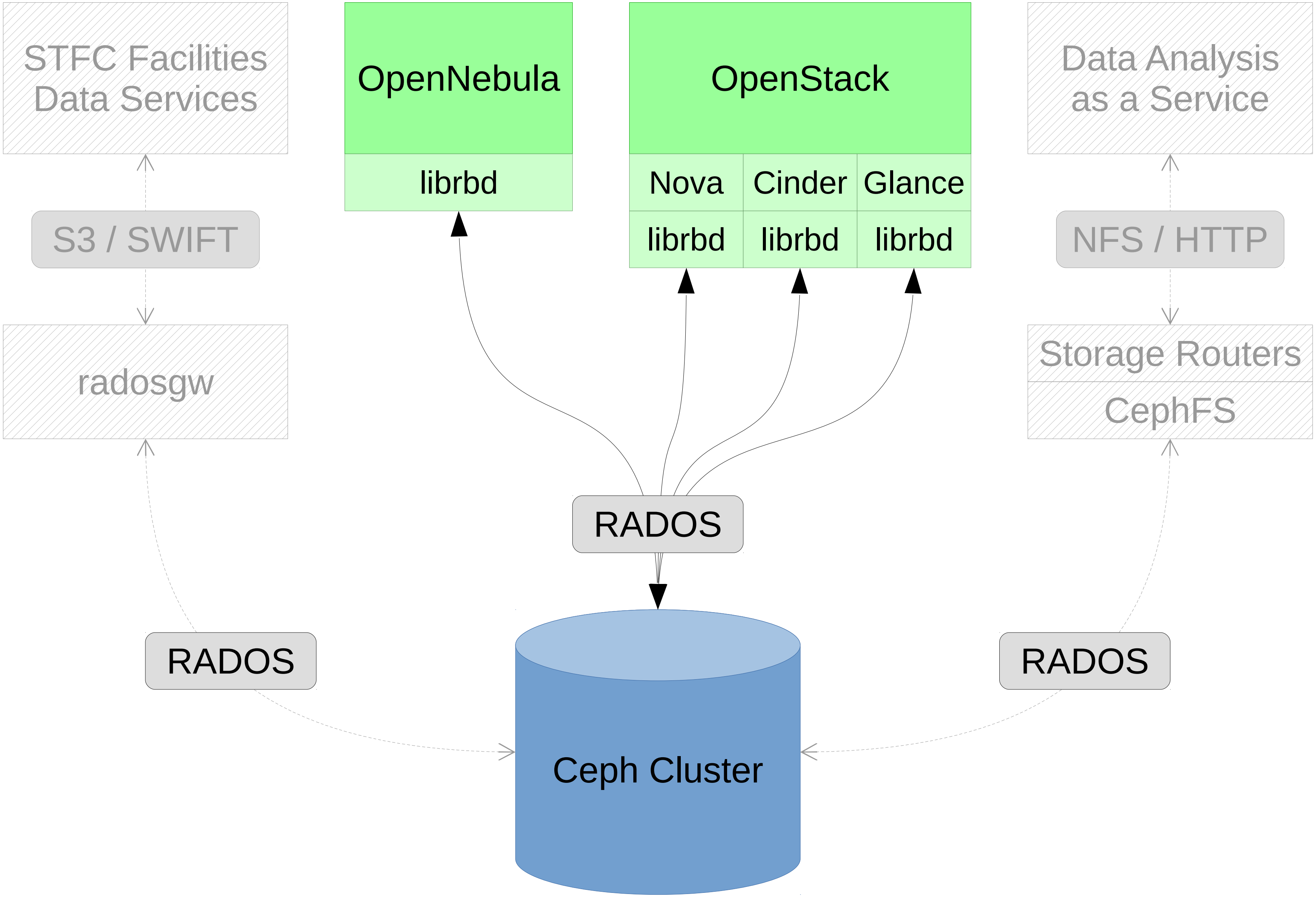
Sirius — Architecture



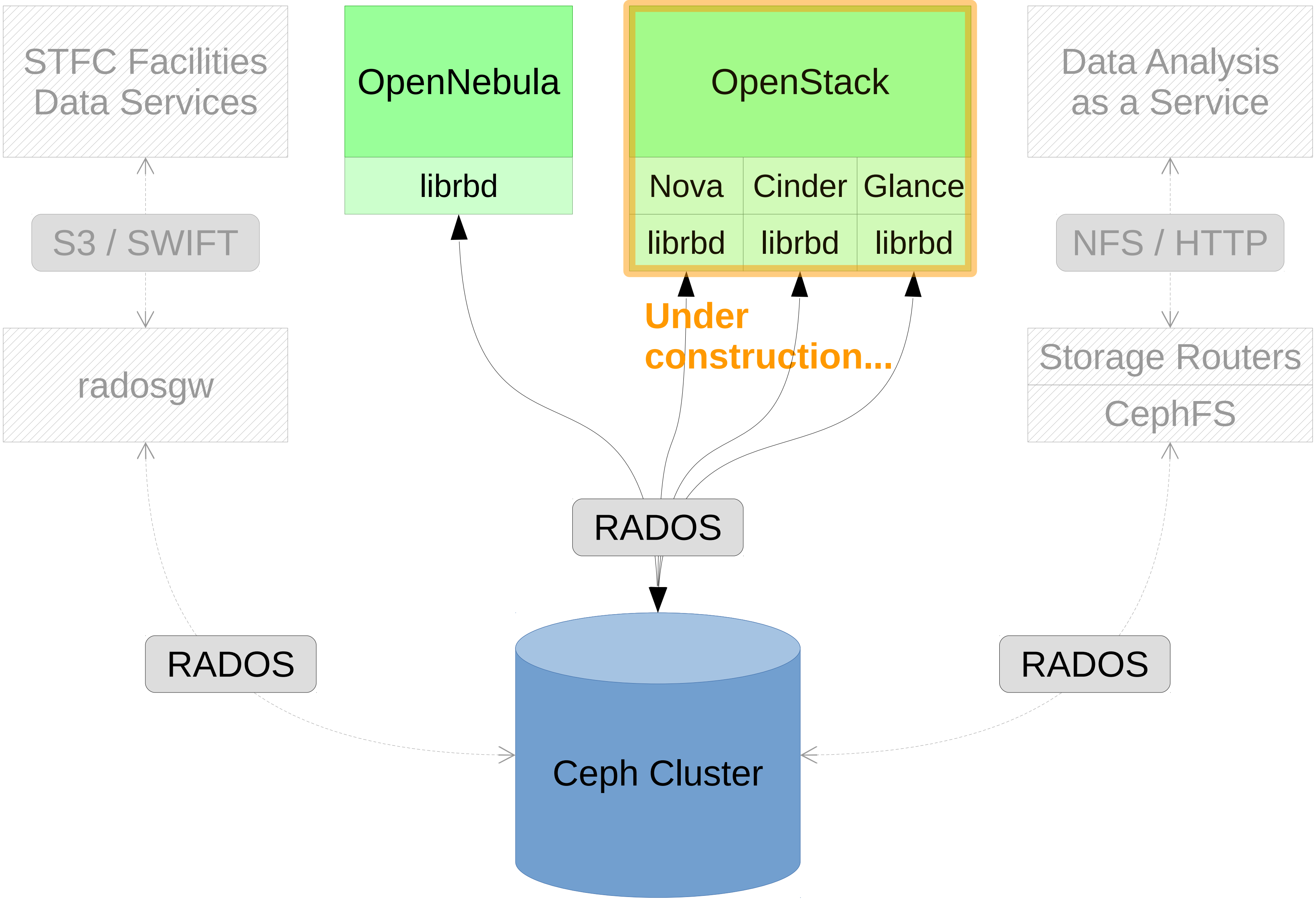
Sirius — Clients



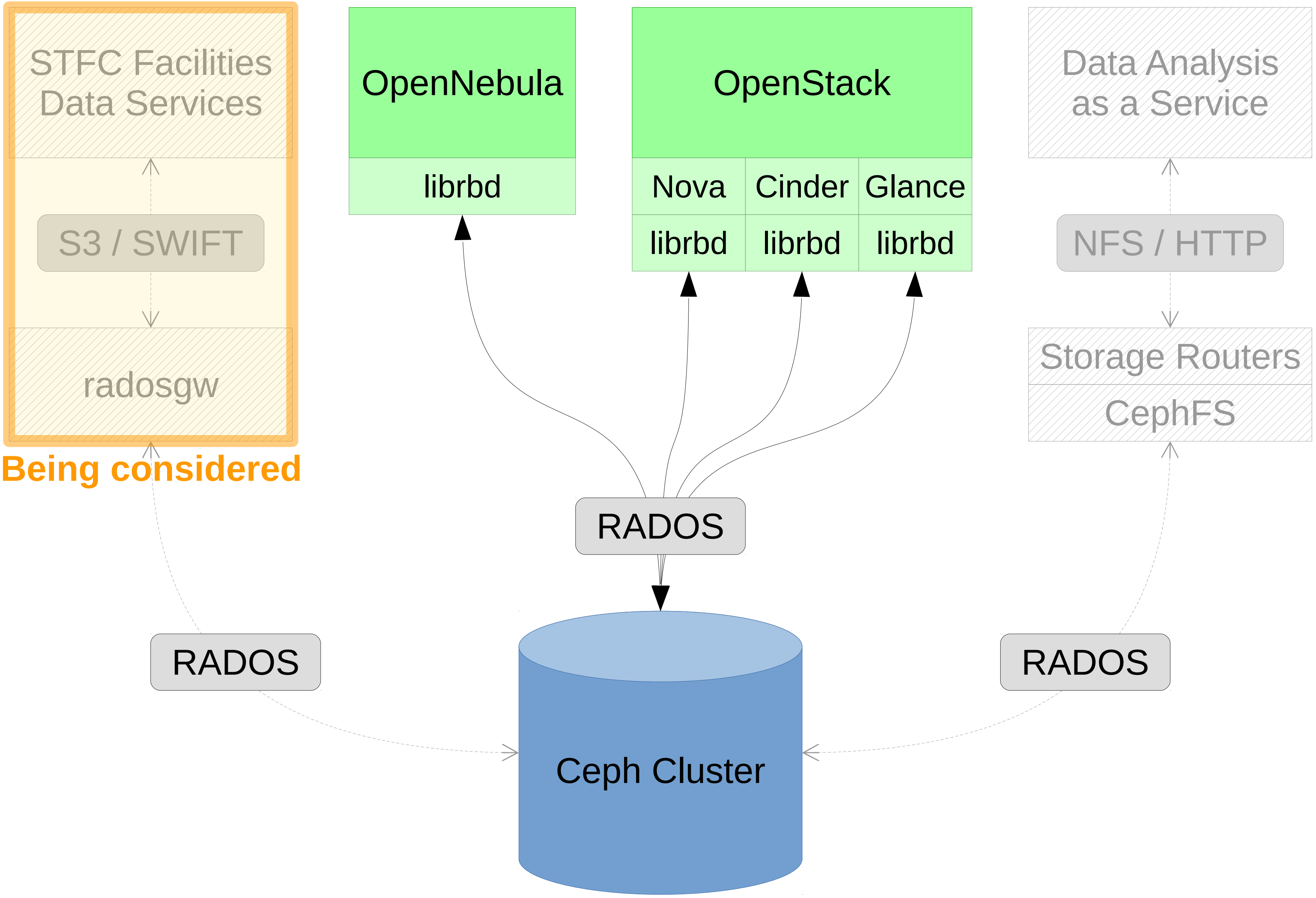
Sirius — Protocols



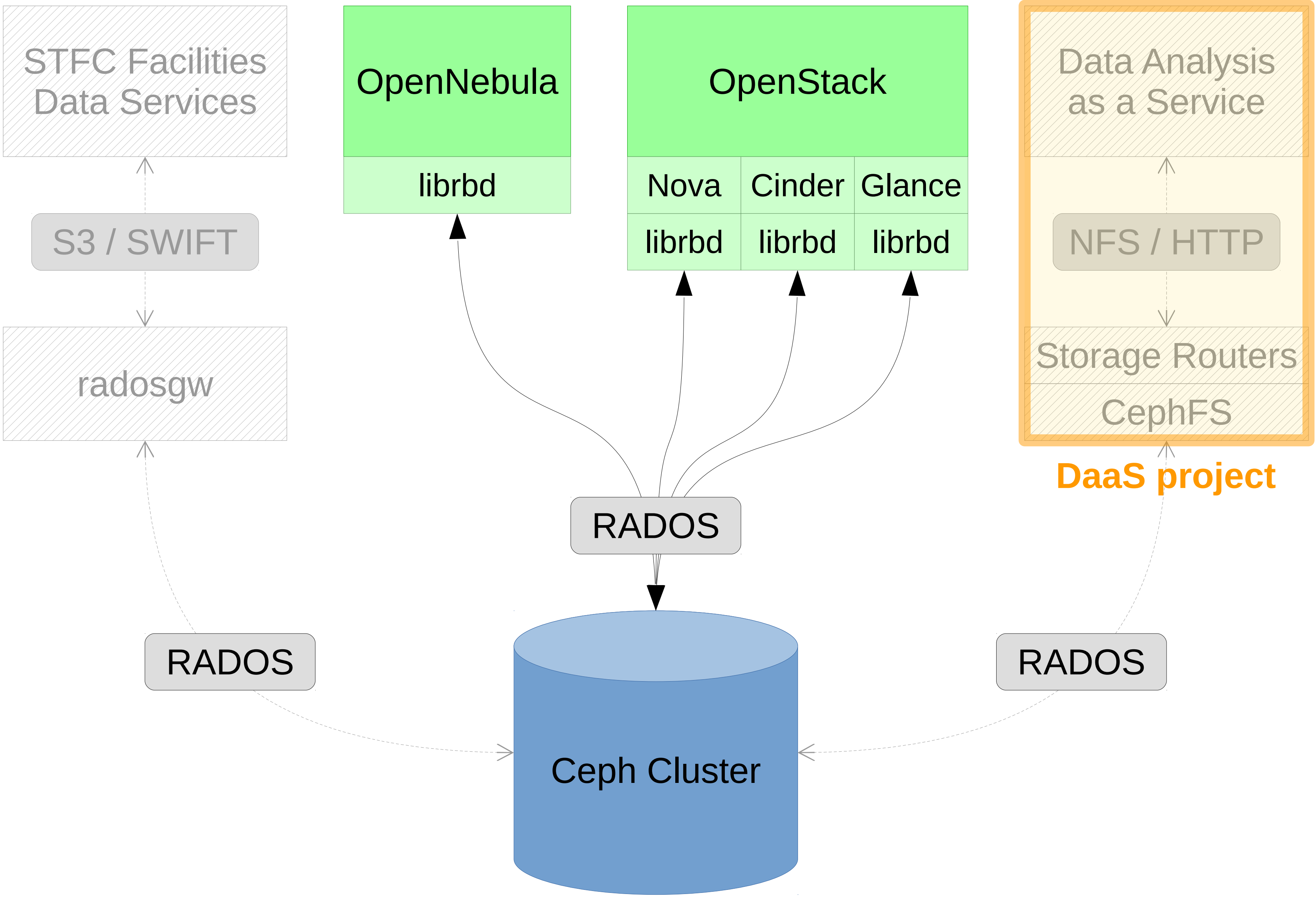
Sirius — Development



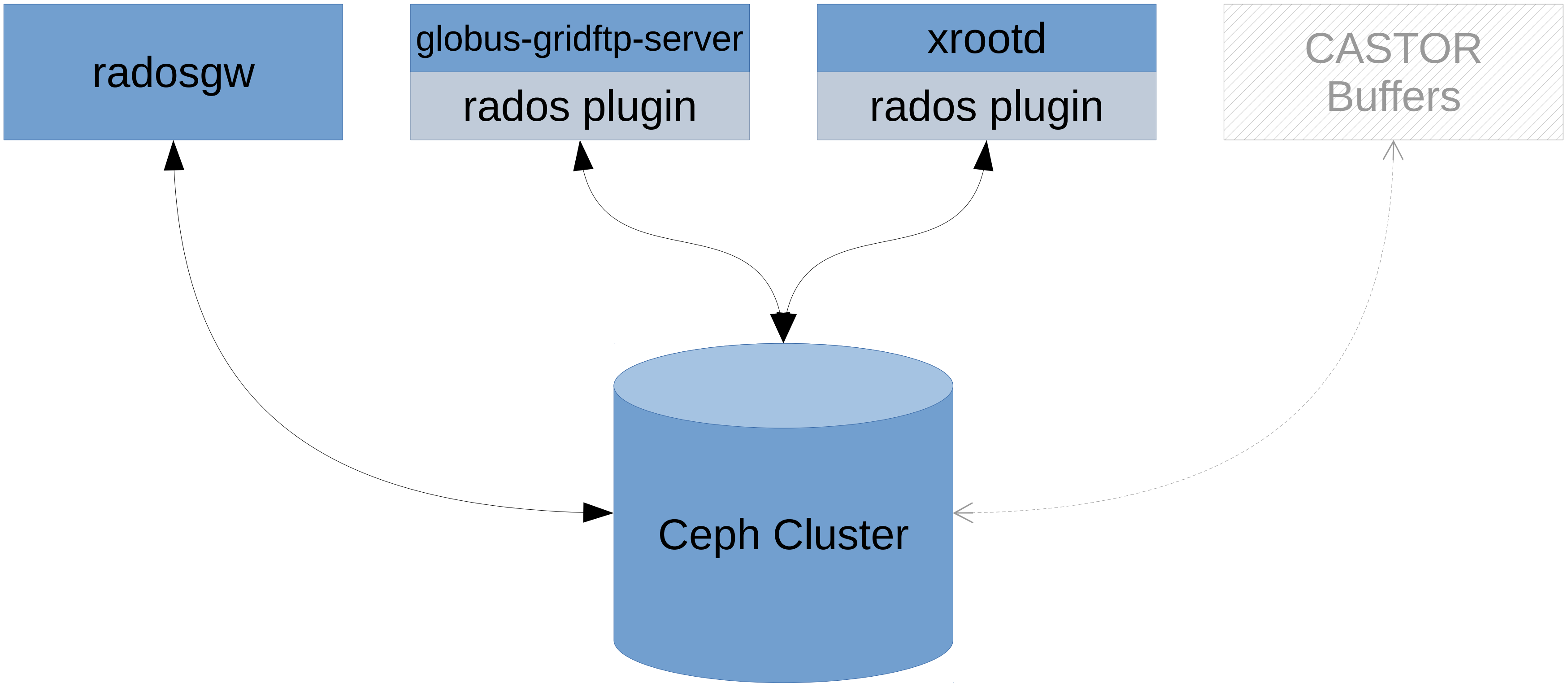
Sirius — Future



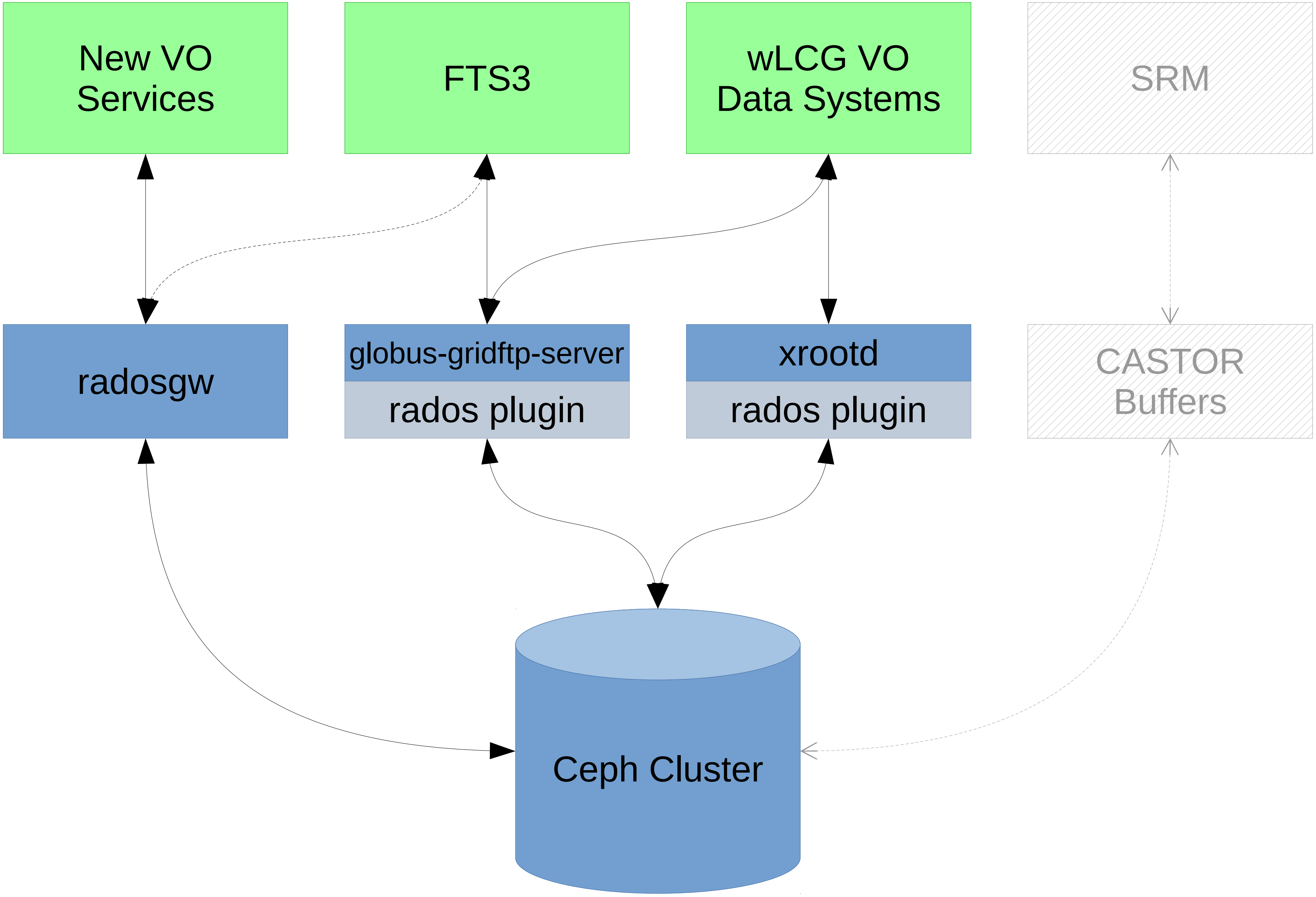
Sirius — Future



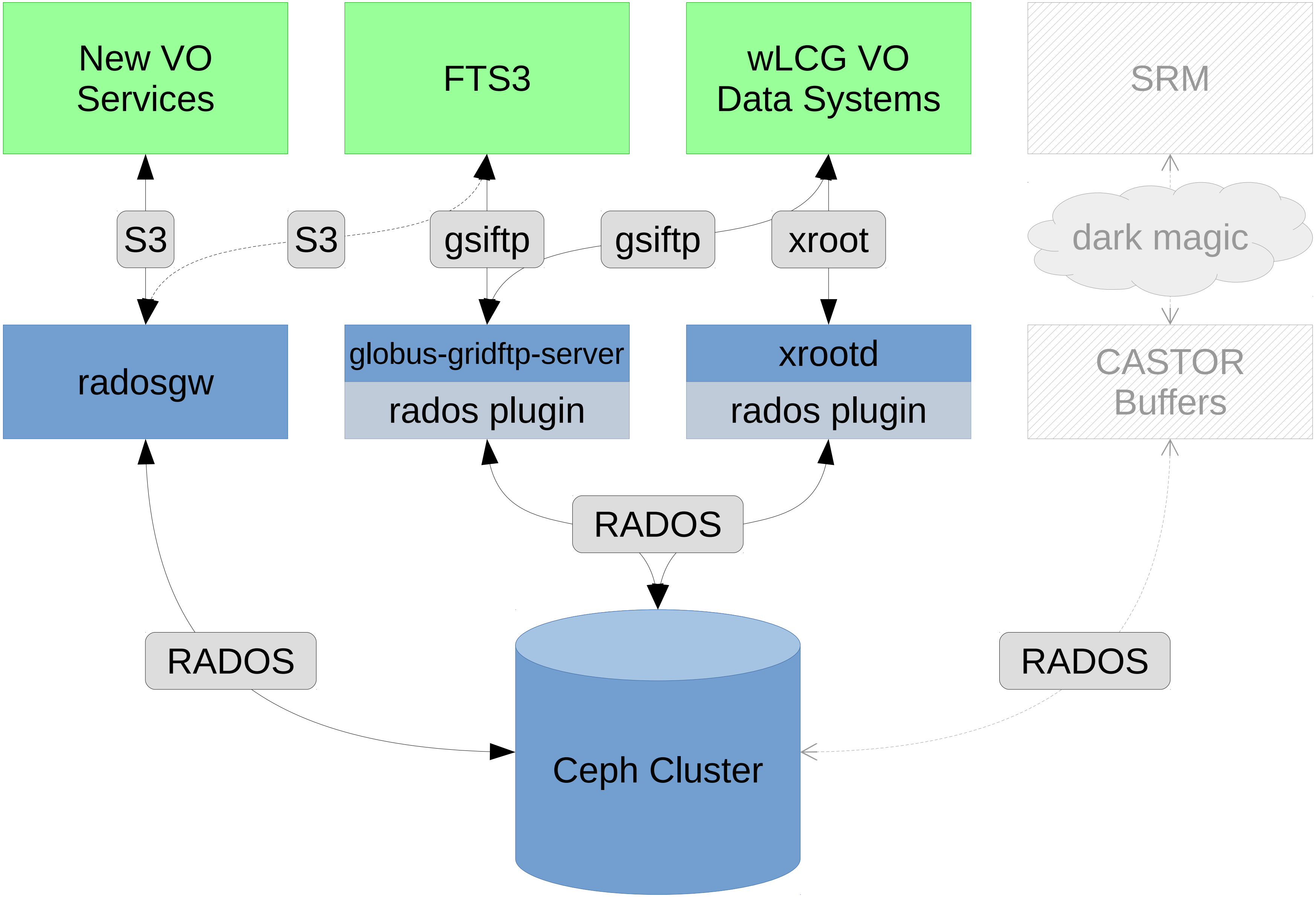
Echo — Architecture



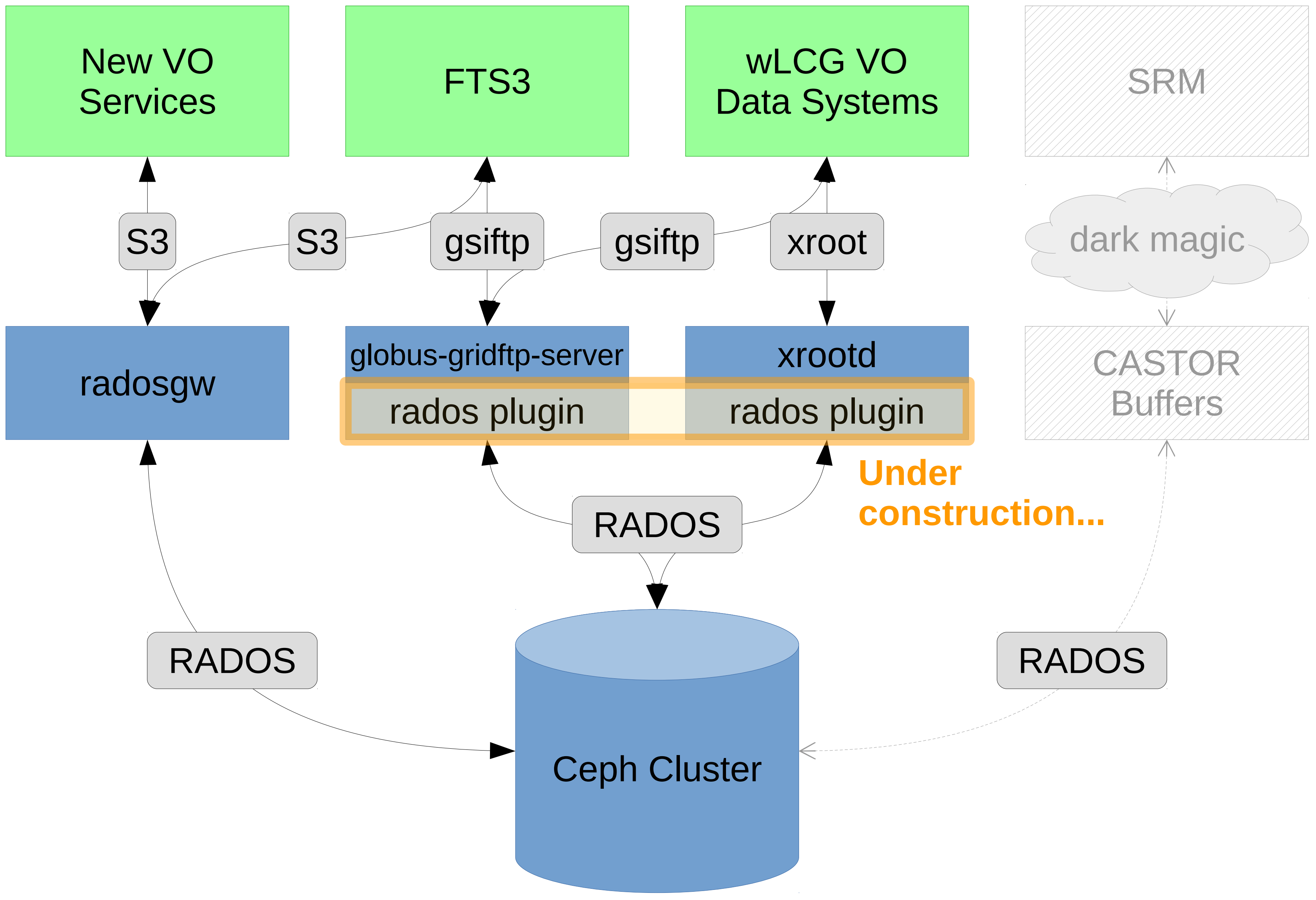
Echo — Clients



Echo — Protocols

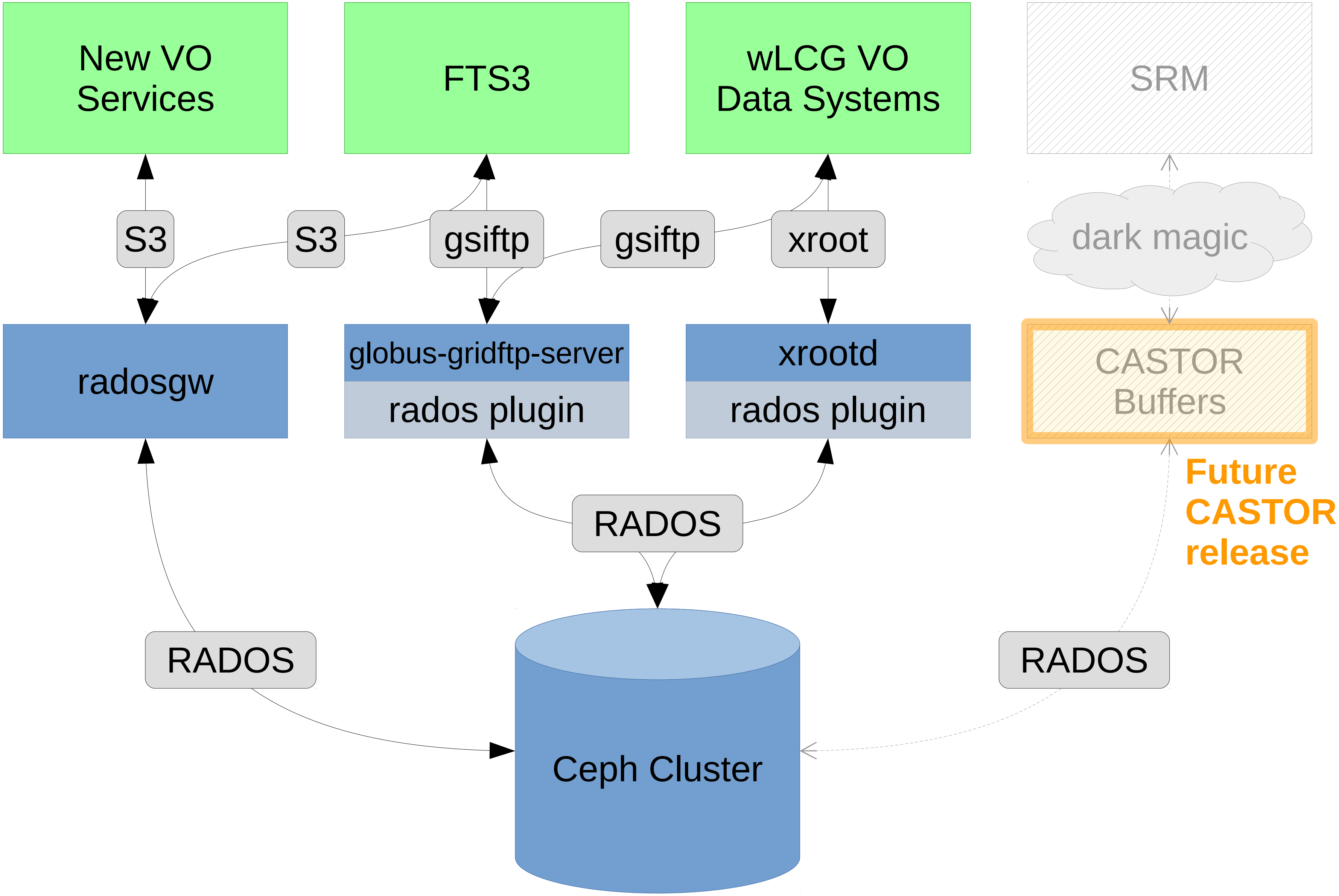


Echo — Development



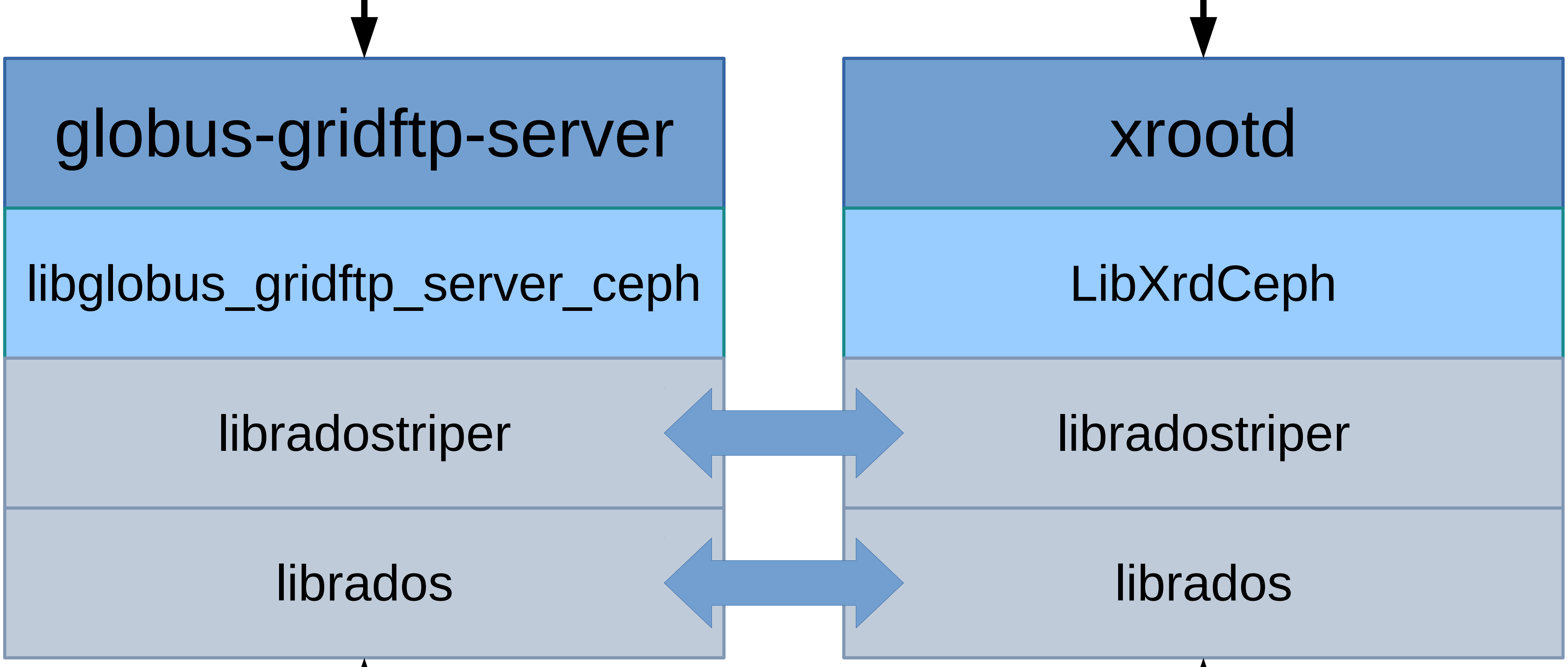
Under construction...

Echo — Future



Future CASTOR release

Echo — Gateways



globus-gridftp-server

libglobus_gridftp_server_ceph

libradostriper

librados

xrootd

LibXrdCeph

libradostriper

librados

Common lower-level libraries mean object format is identical between gateways - both can access same objects.

Echo — Gateways

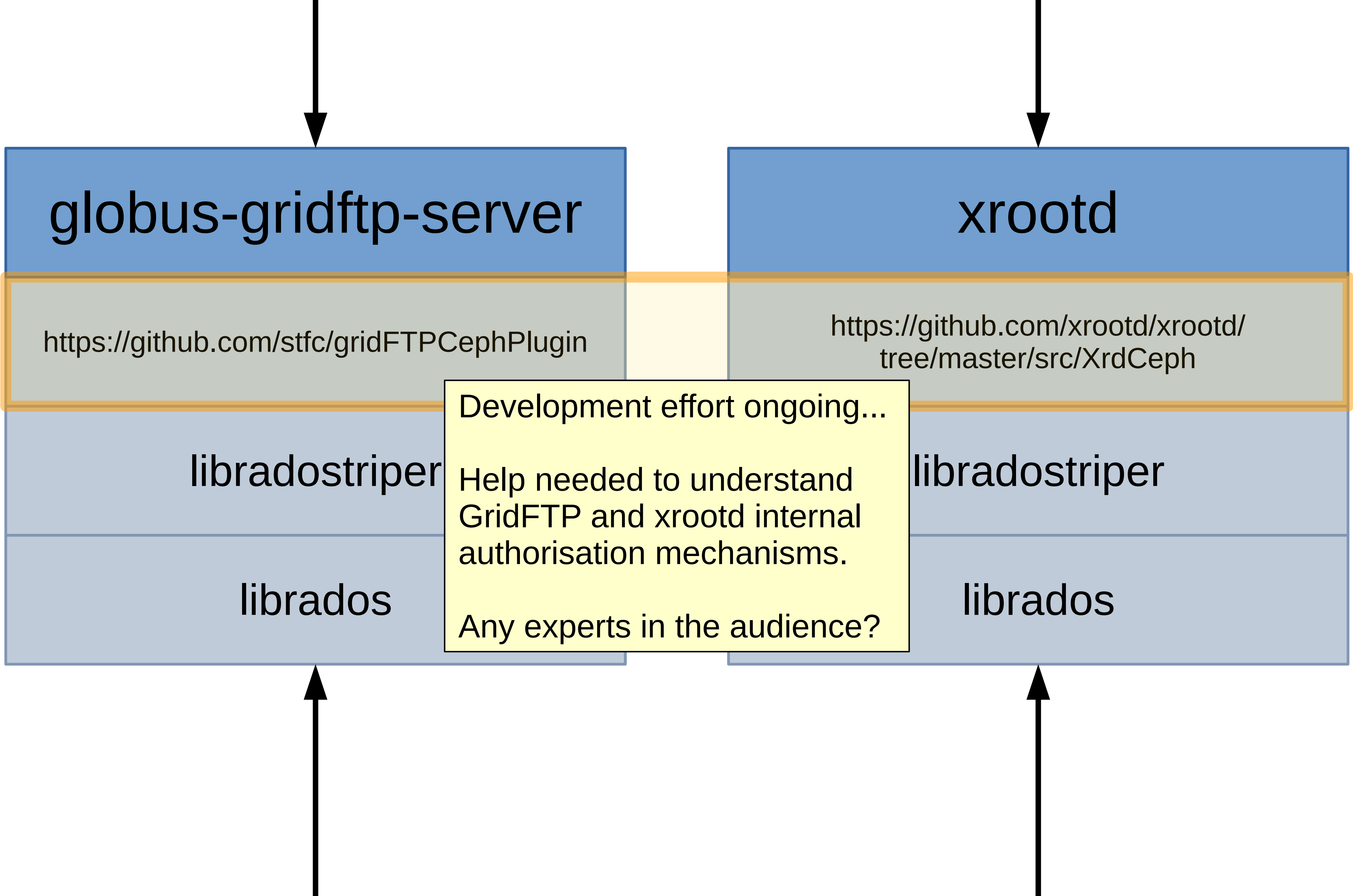
What works?

- X509 authentication
- Access same data with GridFTP and xrootd
- Throughput high enough
  - Reached line-rate
  - Except when FTS resets chunk-size its.cern.ch/jira/browse/FTS-521

What doesn’t work?

- Authorisation
  - Any authenticated user can read/write to any pool
- Accessing data written by GridFTP/xrootd via S3
  - RADOSGW pre-dates libradosstriper, underlying object format different

Echo — Gateways


globus-gridftp-server
https://github.com/stfc/gridFTPCephPlugin
libradostriper
librados
xrootd
https://github.com/xrootd/xrootd/tree/master/src/XrdCeph
libradostriper
librados

Development effort ongoing...

Help needed to understand GridFTP and xrootd internal authorisation mechanisms.

Any experts in the audience?

Echo — RADOS Gateways

- Discussing usage of S3 (and Swift) with Vos
  - ATLAS
    - Keen! Start writing log files soon (~9% CASTOR load)
  - CMS
    - Playing with S3, but no appetite to change yet
  - LHCb
    - Interested, but motivation not high
    - DIRAC developers very keen to support S3
  - Alice
    - No desire to use S3
    - Require a specialised XrootD configuration – no support

Where next?

- Upgrade Sirius and Echo
  - From SL6 to SL7
  - Ceph to Jewel (LTS)
- Echo development effort focused on interfaces
  - Request authorisation very high priority
- 5PiB Echo demonstrator by July 1st
  - Production deployment decision by October

Thanks!

Questions?