

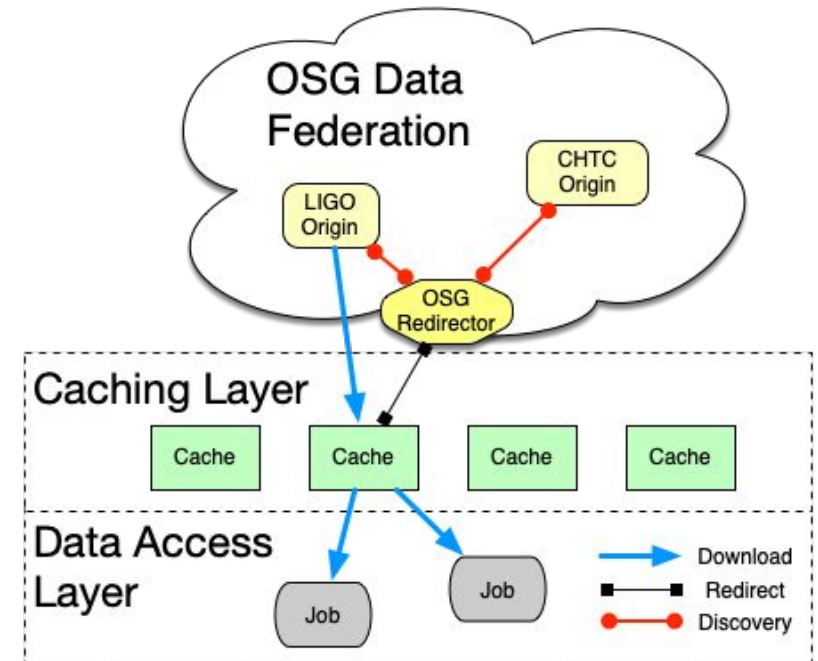


Open Science Data Federation - OSDF

operation and monitoring

Open Science Data Federation

- Origin: Storage of the data from multiple Virtual Organizations (VO);
- Redirector: Process the data request to direct the request to the appropriate origin;
- Cache: Storage to provide data geographically close to the execution points and access points;
- Data access: It is possible to use different tools. Using OSDF commands is possible to fetch the files from the closest cache (GeoIP);



Cache, origins, and redirector based on the XRootD technology:

<https://xrootd.slac.stanford.edu>

<https://pelicanplatform.org>



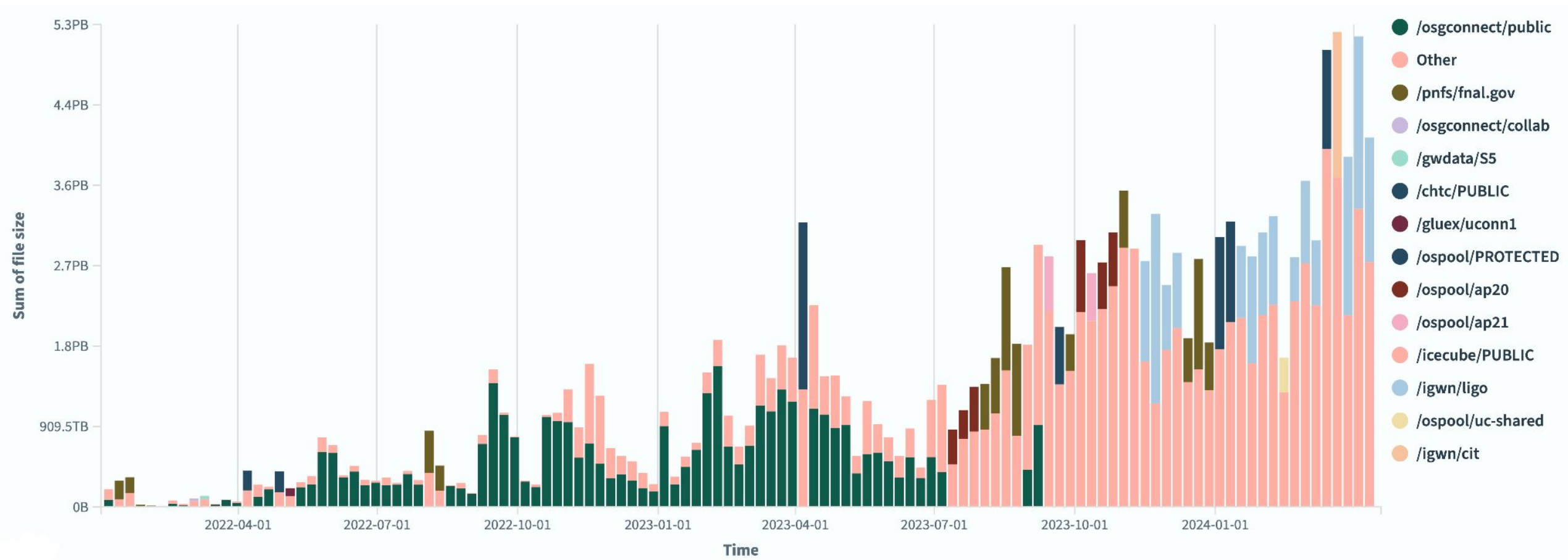
	Origin
	11 Sites, 9 Institutions
	Cache
	29 Sites, 20 Institutions
	Cache and Origin
	6 Sites, 6 Institutions

Metric	Value
Caches	37 across 14 Inst.
Origins	15 across 6 Inst.
Transfers	5,861,721,390
Total size of transfers	294.7PB
Number of checks/monitoring	352
Number of Kubernetes Pods	82

Table 1: OSDF top-level statistics, 2022-2023.

Project	Files request
LIGO	6,893,702,577
LIGO user-specific	565,677,512
fnal.gov - Nova	492,252,998
fnal.gov - Minerva	435,978,085
fnal.gov - Dune	412,684,786
OSG Collaborations	226,827,338
fnal.gov - uboone	156,274,029
LIGO gwdata - O3a - uboone	145,505,762
LIGO gwdata - O2 - uboone	103,851,076
OSG user	98,685,798
Other	766,798,555

Table 2: Files requested by projects between 2018 and 2023.




```
fabioandrijauskas@macf ~ % wget https://osdf-director.osg-htc.org/nrp/osdf/bbsolar/1mb
--2024-10-02 09:07:44-- https://osdf-director.osg-htc.org/nrp/osdf/bbsolar/1mb
Resolving osdf-director.osg-htc.org (osdf-director.osg-htc.org)... 128.105.82.132
Connecting to osdf-director.osg-htc.org (osdf-director.osg-htc.org)|128.105.82.132|:443... c
onnected.
HTTP request sent, awaiting response... 307 Temporary Redirect
Location: https://fdp-d3d-cache.nationalresearchplatform.org:8443/nrp/osdf/bbsolar/1mb [foll
owing]
--2024-10-02 09:07:45-- https://fdp-d3d-cache.nationalresearchplatform.org:8443/nrp/osdf/bb
solar/1mb
Resolving fdp-d3d-cache.nationalresearchplatform.org (fdp-d3d-cache.nationalresearchplatf
orm.org)... 132.249.252.218
Connecting to fdp-d3d-cache.nationalresearchplatform.org (fdp-d3d-cache.nationalresearchplat
form.org)|132.249.252.218|:8443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 1000000 (977K)
Saving to: '1mb.18'

1mb.18          100%[=====>] 976.56K  --.-KB/s    in 0.1s

2024-10-02 09:07:45 (9.51 MB/s) - '1mb.18' saved [1000000/1000000]
```

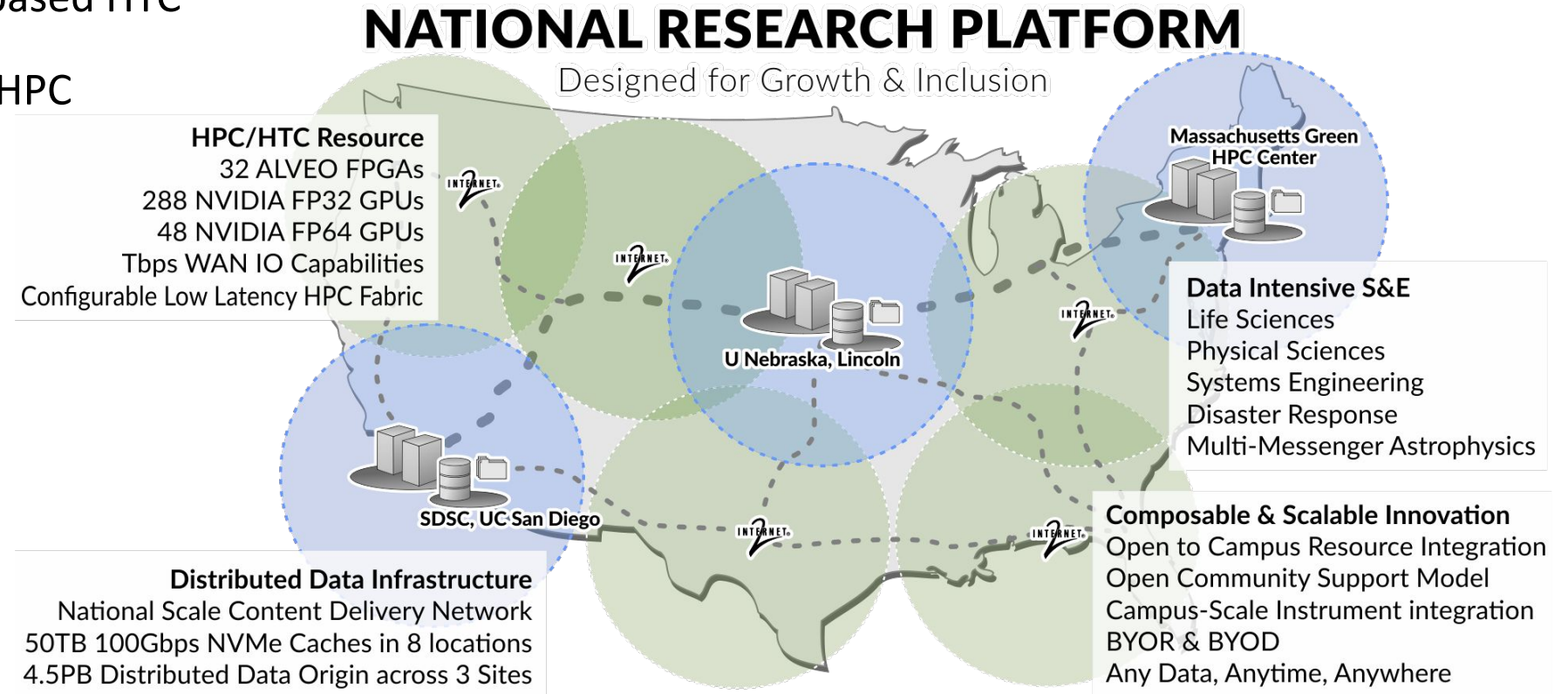
STATE	HOST	ICONS	OK	WA	UN	CR	PD
UP	ap1-origin.facility.path-cc.io	≡	2	0	0	0	0
UP	ap20.uc.osg-htc.org	≡	6	0	0	0	0
UP	ap21.uc.osg-htc.org	≡	5	1	0	0	0
UP	ap22.uc.osg-htc.org	≡	6	0	0	0	0
UP	ap23.uc.osg-htc.org	≡	6	0	0	0	0
UP	ap40.uw.osg-htc.org	≡	2	2	0	0	0
UP	cn447.storrs.hpc.uconn.edu	≡	2	0	0	0	0
UP	et-origin.cism.ucl.ac.be	≡	2	0	0	0	0
UP	fdp-d3d-origin.nationalresearchplatform.org	≡	1	0	0	0	0
UP	fiona.its.hawaii.edu	≡	1	0	0	0	0
UP	gandalf.phys.uconn.edu	≡	2	0	0	0	0
UP	hcc-stashcache-origin.unl.edu	≡	3	0	0	0	0
UP	ingrid-se09.cism.ucl.ac.be	≡	2	0	0	0	0
UP	kagra-dsr-b2.icrr.u-tokyo.ac.jp	≡	1	0	0	0	0
UP	koastore-osdf.its.hawaii.edu	≡	1	0	0	0	0
UP	mghpcc-origin.nationalresearchplatform.org	≡	2	0	0	0	0
UP	mo-origin.nationalresearchplatform.org	≡	0	1	0	0	0
UP	ncar.nationalresearchplatform.org	≡	2	0	0	0	0
UP	nrdstor.nationalresearchplatform.org	≡	1	0	0	0	0
UP	origin-auth2000.chtc.wisc.edu	≡	2	0	0	0	0
UP	origin-ifo.ligo.caltech.edu	≡	2	0	0	0	0
UP	origin-shared.ligo.caltech.edu	≡	2	0	0	0	0
UP	origin-staging.ligo.caltech.edu	≡	2	0	0	0	0
UP	origin.ligo.caltech.edu	≡	2	0	0	0	0
UP	osdforigin.fnal.gov	≡	2	0	0	0	0
UP	osdfest.t2.ucsd.edu	≡	1	0	0	0	0
UP	osg-sci1.datacollaboratory.org	≡	1	0	0	0	0
UP	pelican-osdf-public.tempest.uchicago.edu	≡	1	0	0	0	0
UP	prp01.ifa.hawaii.edu	≡	1	0	0	0	0
UP	purdue-origin.nationalresearchplatform.org	≡	1	0	0	0	0
UP	sc-origin2000.chtc.wisc.edu	≡	2	0	0	0	0
UP	sc-origin.chtc.wisc.edu	≡	1	0	0	0	0
UP	sci-xrootd.jlab.org	≡	2	0	0	0	0
UP	sdsc-origin.nationalresearchplatform.org	≡	2	0	0	0	0
UP	sdsc-s3-origin.nationalresearchplatform.org	≡	1	0	0	0	0
UP	sdsc-xenon-origin.nationalresearchplatform.org	≡	1	0	0	0	0
UP	se1.oscer.ou.edu	≡	2	0	0	0	0
UP	stash-origin.icecube.wisc.edu	≡	2	0	0	0	0
UP	xrootd-local.unl.edu	≡	1	0	0	0	0

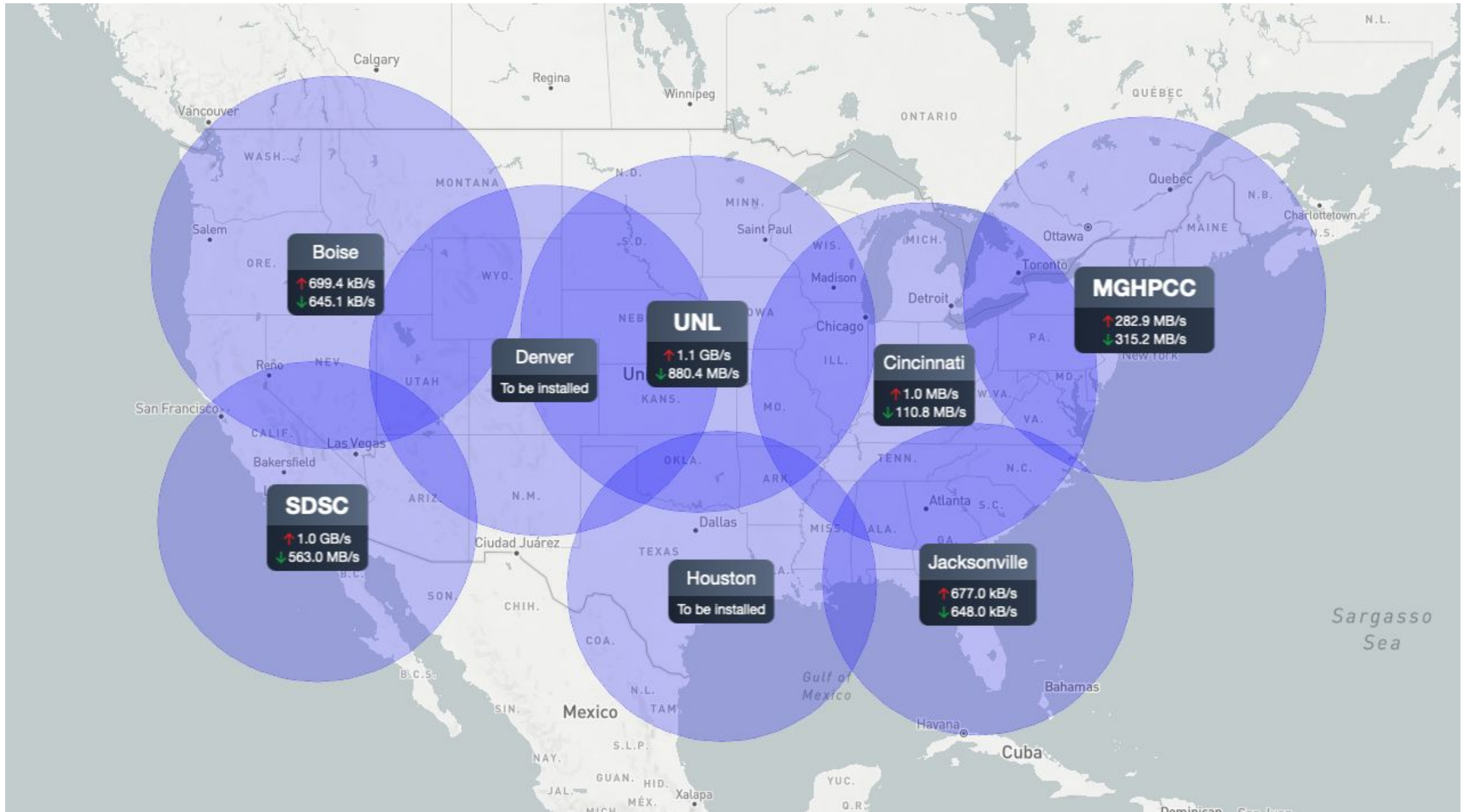
STATE	HOST	ICONS	OK	WA	UN	CR	PD
UP	amst-fiona.nationalresearchplatform.org	≡	6	0	0	0	0
UP	amst-osdf-xcache01.es.net	≡	6	0	0	0	0
UP	ccstashcachell01.in2p3.fr	≡	5	0	0	0	0
UP	cf-ac-uk-cache.nationalresearchplatform.org	≡	6	0	0	0	0
UP	daejeon-kreonet-net.nationalresearchplatform.org	≡	7	0	0	0	0
UP	ds-914.cr.cnaf.infn.it	≡	4	0	0	0	0
UP	dtn-pas.bois.nrp.internet2.edu	≡	6	0	0	0	0
UP	dtn-pas.cinc.nrp.internet2.edu	≡	6	0	0	0	0
UP	dtn-pas.hous.nrp.internet2.edu	≡	6	0	0	0	0
UP	dtn-pas.jack.nrp.internet2.edu	≡	7	0	0	0	0
UP	fdp-d3d-cache.nationalresearchplatform.org	≡	5	0	0	0	0
UP	its-condor-xrootd1.syr.edu	≡	5	0	0	0	0
UP	ligo.hpc.swin.edu.au	≡	4	0	0	0	0
UP	lond-osdf-xcache01.es.net	≡	6	0	0	0	0
UP	mghpcc-cache.nationalresearchplatform.org	≡	6	0	0	0	0
UP	osdf-cache.sprace.org.br	≡	4	0	0	0	0
DOWN	osdf-uw-cache.svc.osg-htc.org	≡	2	0	0	4	0
UP	osg-chicago-stashcache.nrp.internet2.edu	≡	7	0	0	0	0
UP	osg-gftp2.pace.gatech.edu	≡	5	0	0	0	0
UP	osg-hawk-cache.cc.lehigh.edu	≡	4	0	0	0	0
UP	osg-houston-stashcache.nrp.internet2.edu	≡	7	0	0	0	0
UP	osg-new-york-stashcache.nrp.internet2.edu	≡	7	0	0	0	0
UP	osg-stash-sfu-computecanada-ca.nationalresearchplatform.org	≡	5	0	0	0	0
UP	osg-sunnyvale-stashcache.nrp.internet2.edu	≡	6	0	0	0	0
UP	sdsc-cache.nationalresearchplatform.org	≡	7	0	0	0	0
UP	singapore.nationalresearchplatform.org	≡	7	0	0	0	0
UP	stash.farm.particle.cz	≡	3	1	0	0	0
UP	stashcache-edi-scotgrid-ac-uk.nationalresearchplatform.org	≡	5	0	0	0	0
UP	stashcache.jinr.ru	≡	5	0	0	0	0
UP	stashcache.ligo.caltech.edu	≡	4	0	0	0	0
UP	stashcache.t2.ucsd.edu	≡	6	0	0	0	0
UP	unl-cache.nationalresearchplatform.org	≡	6	0	0	0	0
UP	xcachevirgo.pic.es	≡	5	0	0	0	0

- 1 Cache (50TB for cache) and one origin (1.6PB for origin storage): San Diego Super Computer Center - San Diego - California.
- 1 Cache (42TB for cache) and one origin (1.2 PB for origin storage): University of Nebraska-Lincoln - Lincoln - Nebraska.
- 1 Cache (29TB for cache) and one origin (1.2PB for origin storage): Massachusetts Green High-Performance Computing Center - Holyoke - Massachusetts.
- 1 Cache: Internet2 - Boise - Idaho (42TB for cache).
- 1 Cache: Internet2 - Jacksonville - Florida (42TB for cache).
- 1 Cache: Internet2 - Denver - Colorado (42TB for cache).
- 1 Cache: Northeastern University - Boston - Massachusetts (in installation).
- 1 cache (50 TB) and one origin (1.6TB) SDSC-UCSD. - 1 cache (42 TB) and one origin (1.2 PB) UNL -
- 1 cache (29 TB) and one origin (1.2PB) MGHPCC -
- Internet2 hubs: Boise - Idaho (42 TB), Houston - Texas (in progress), Jacksonville - Florida (42 TB), Denver - Colorado (42 TB), and Northeastern University Boston. Existing Internet2 caches in Chicago, New York, and Kansas City. -
- 1 cache (7 TB) and one origin (7 TB): The University of Tokyo.
- 1 cache (14 TB): The National Center for Atmospheric Research (NCAR), Boulder
- 1 origin (337 TB): University of Nebraska-Lincoln

System architecture

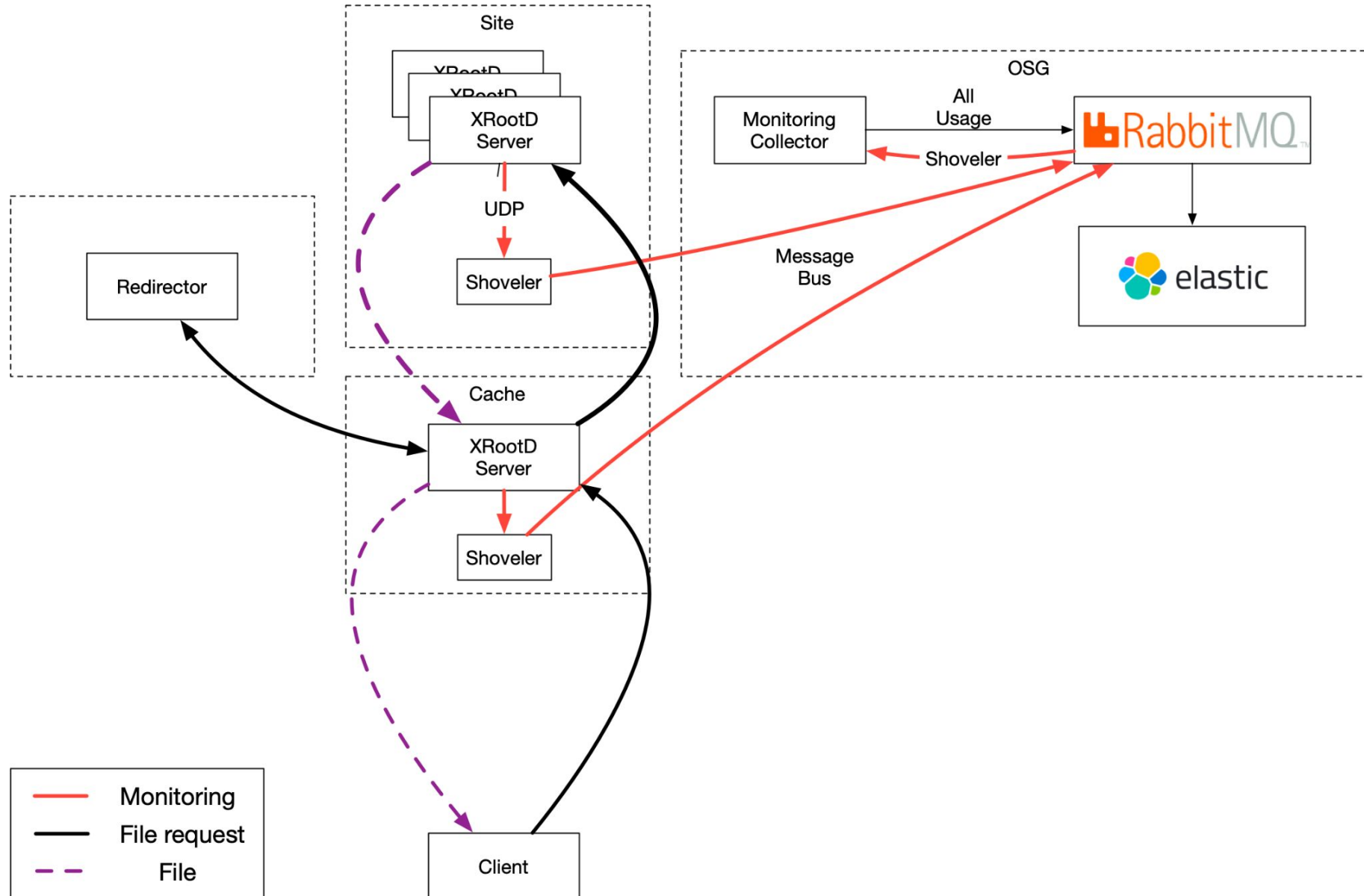
- A distributed system spanning multiple locations
- Three semi-independent components:
 - FP32-optimized GPU-based HTC
 - GPU and FPGA-based HPC
 - Storage
- All managed through a unified Kubernetes setup





<https://nrp-website.vercel.app>

OSDF



Conclusion and remarks

- OSDF is “free”; let’s talk about staging your data!
- More caches, more data access.
- Several projects - public and private
- Do you have data on your computer under your desk? Let’s share!

fandrijauskas@ucsd.edu