

Collaborations on the OSG Fabric of Sciences

By Pascal Paschos

NSF Ecosystem Lightning Talks

March 2, 2022

CI Compass Cyberinfrastructure for NSF Major
Facilities Workshop

Cyberinfrastructure for NSF Major Facilities Workshop

Collaborations on the OSG Fabric of Services

Pascal Paschos
OSG Area Coordinator for Collaboration support

The OSG Consortium

- OSG is a consortium dedicated to the advancement of all of open science via the practice of distributed High Throughput Computing (dHTC), and the advancement of its state of the art.
- Governed by the OSG Council, maintaining its by-laws, and electing an executive director for 2 year renewable terms to coordinate a program of work.
- OSG is a collaboration between CI, software, and science professionals that provides:
 - Training and Consultation
 - Service provisioning
 - Compute, Storage and Data infrastructure
 - Facilitation in the delivery of data products on the critical path to scientific discovery

Categories of Participation

- **Individual Researchers and small groups**, typically in institutional based research groups
- **Campus Research Support Organizations**
 - IT/CI organizations, trained by OSG and in coordination with OSG integrate local resources with OSG and facilitate training and onboarding of multi-disciplinary campus research
- **Multi-institutional Science Teams** (midsize collaborations)
 - Collaborations between research teams in large number of US campuses and institutions
 - Partnerships with international organizations, campuses and institutions
- **Large scale projects:** US-ATLAS, US-CMS, LIGO, Icecube

Service Requirements & Expectations

Individual research or campus group projects

- Primarily use the Open Science pool and Open Science storage shared by all researchers.
- Use the OSG data federation which delivers data in proximity to compute resources.

Collaboration projects

- May contribute dedicated resources forming computing pools under their own control.
 - Some institutional resources allow access by multiple collaborations but not the general science community.
 - Some collaborations use the OS pool and related services in addition to their own pool.
- Data may have access restrictions.
- May contribute infrastructure to the OSG data federation
- May include integration of storage at major international facilities

**Collaborations require a richer set of services
than individual researchers**

Support models

Individual research or campus group projects

- Mainly supported out of OSG research facilitation support team in collaboration with campus professionals

Collaboration projects

- Midsize collaborations typically have minimal or no support teams in integrating with OSG and in management of their workflows and infrastructure. OSG provides such support to **promote** growth in self-reliance and confidence.
- Collaborations with dedicated support teams, typically in major facilities, **partner** with OSG in a shared effort to **facilitate** resolution of issues, support software and data products and infrastructure growth

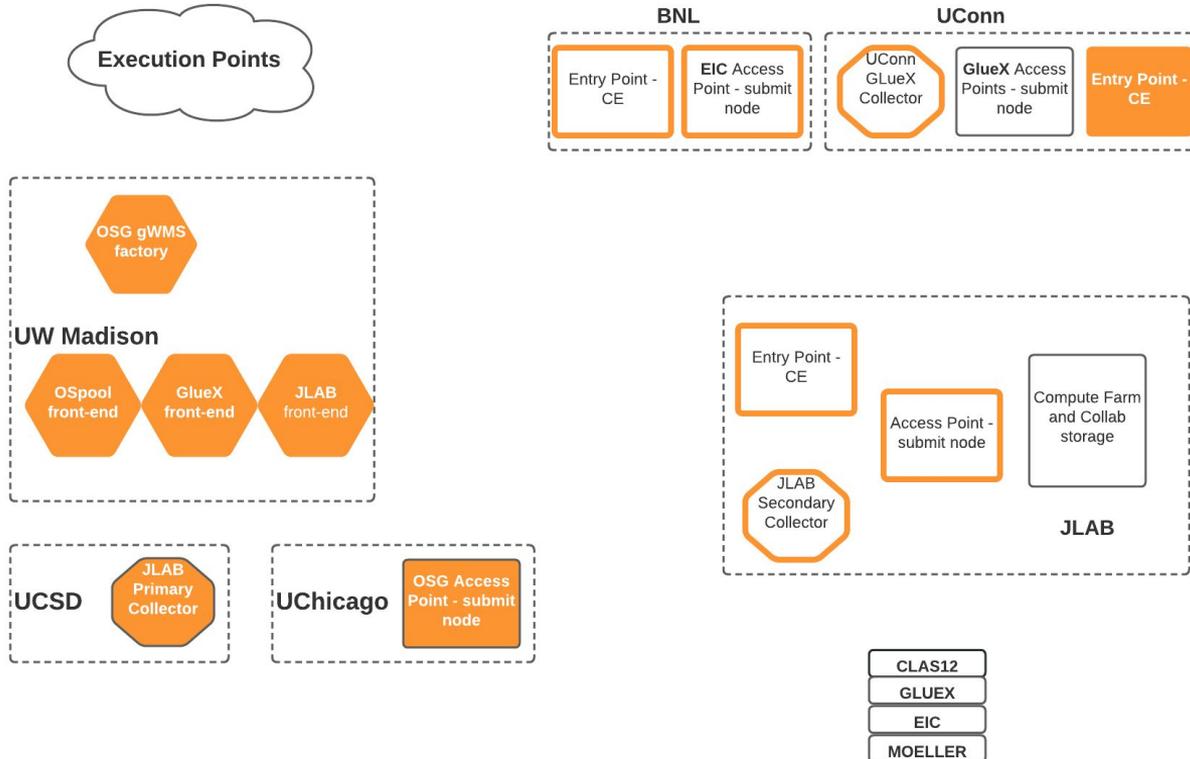
Collaboration Support on OSG

- The Collaboration Support team **coordinates** and **facilitates** service delivery in
 - Onboarding new multi-institutional effort, fostering nascent efforts in exploring the Open Science Fabric of Services and planning
 - Compute, Storage, Network Infrastructure and Hosting Services
 - Scientific Workflows and Data management
 - User and Site support, consultation & documentation
 - Dashboards and monitoring
- We engage via ticketing systems and regularly scheduled calls/virtual meetings, contribute to technical publications and to PATh NSF reports
- At present 13 collaborations are supported by OSG which includes 2 large scale projects. Additional collaborations at Fermilab (FNAL) are supported by FNAL staff in coordination with OSG.

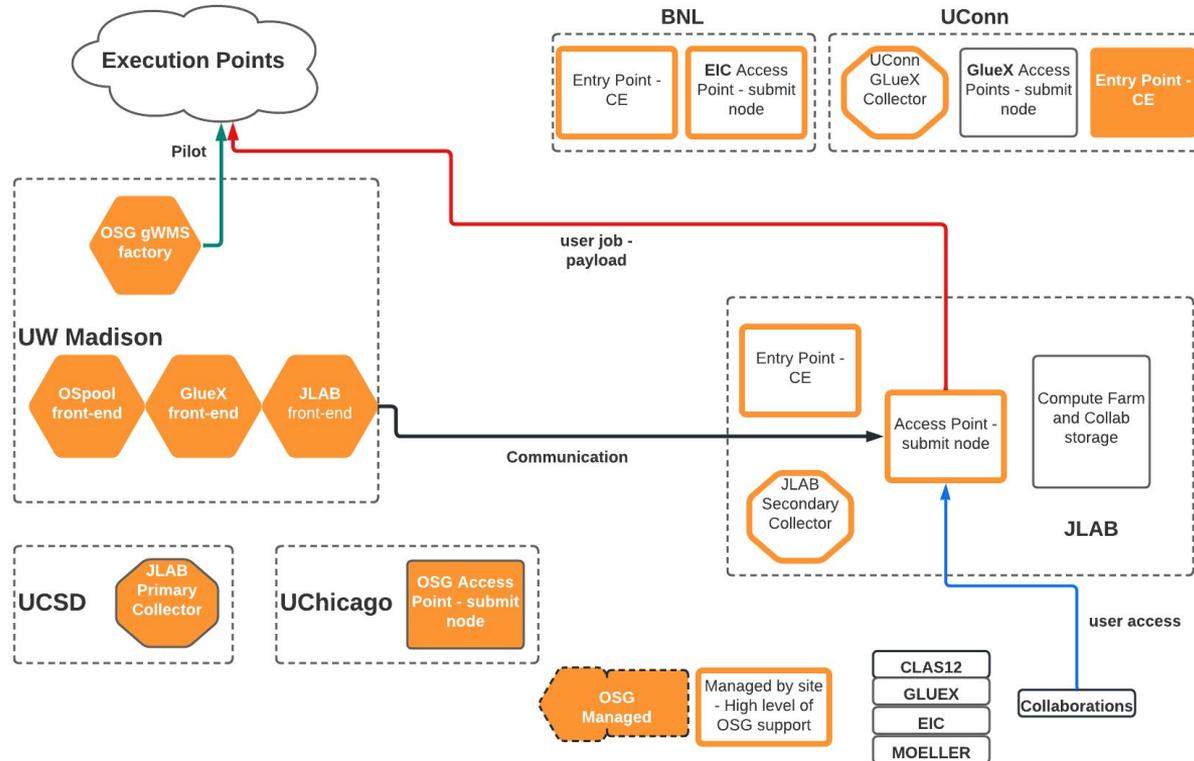
Examples of Collaboration support activities

- Accounting and facilitation of access to OSG-managed infrastructure - OSG Connect
- Set up and consult on access points to OSG compute and/or data federation at facilities that provide dedicated support for the collaboration
- Integrate dedicated collaboration resources into OSG compute and data federations.
 - Deploy & operate caching infrastructure at dedicated collaboration resources
- Support the frontier shaping effort for the HEP community - Snowmass
- Operate data & compute infrastructure that connects components in the end-to-end delivery of science workload
 - This can be quite **complex** in facilities that support multiple collaborations

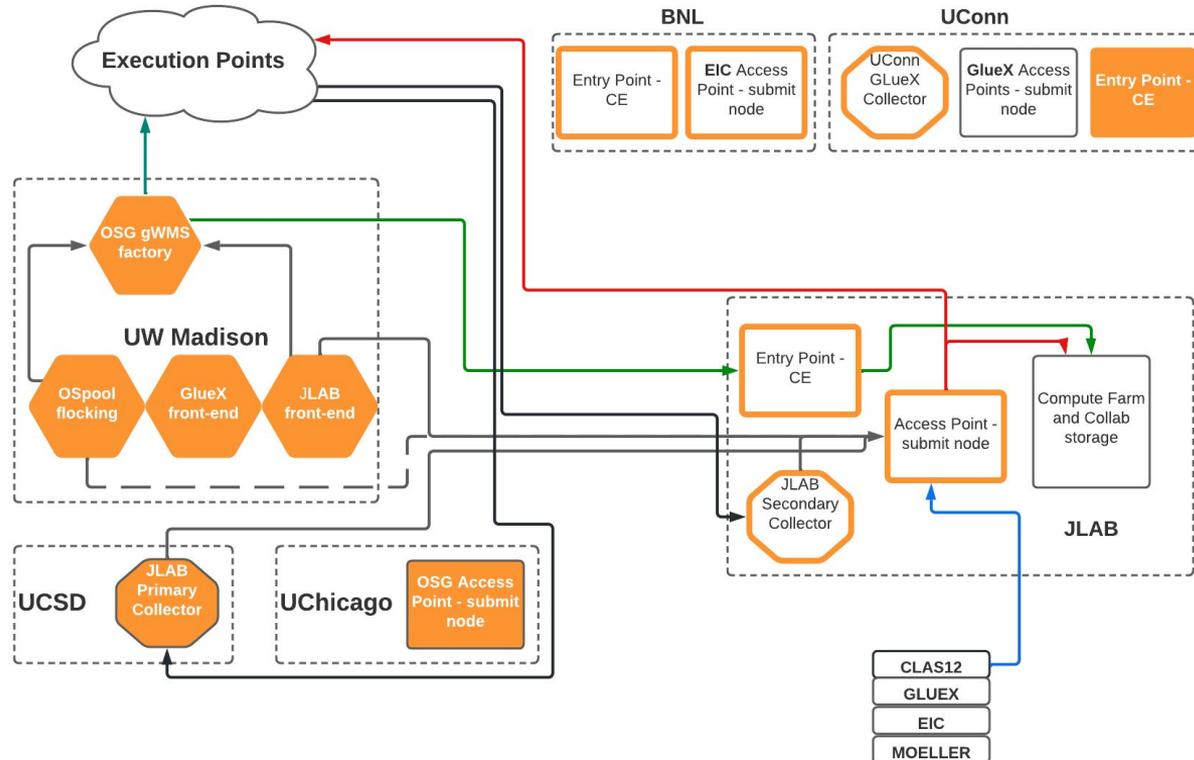
JLab Experiments on the OSG



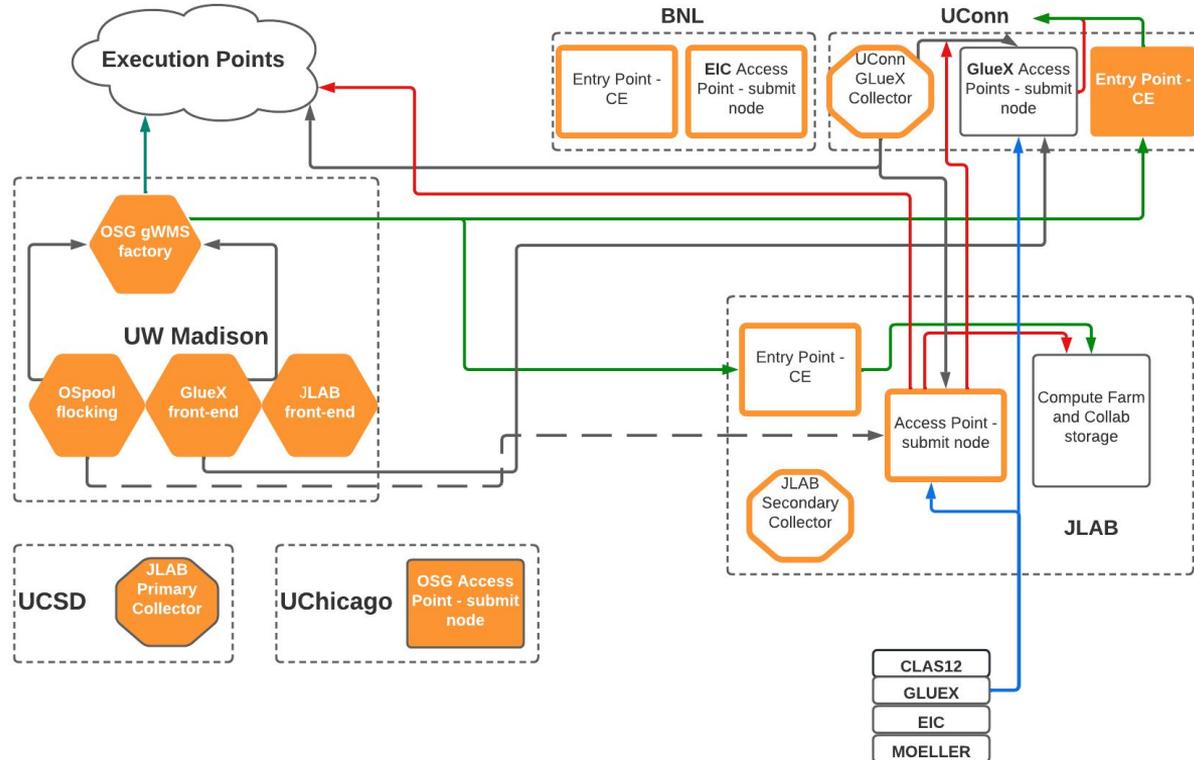
JLab Experiments on the OSG



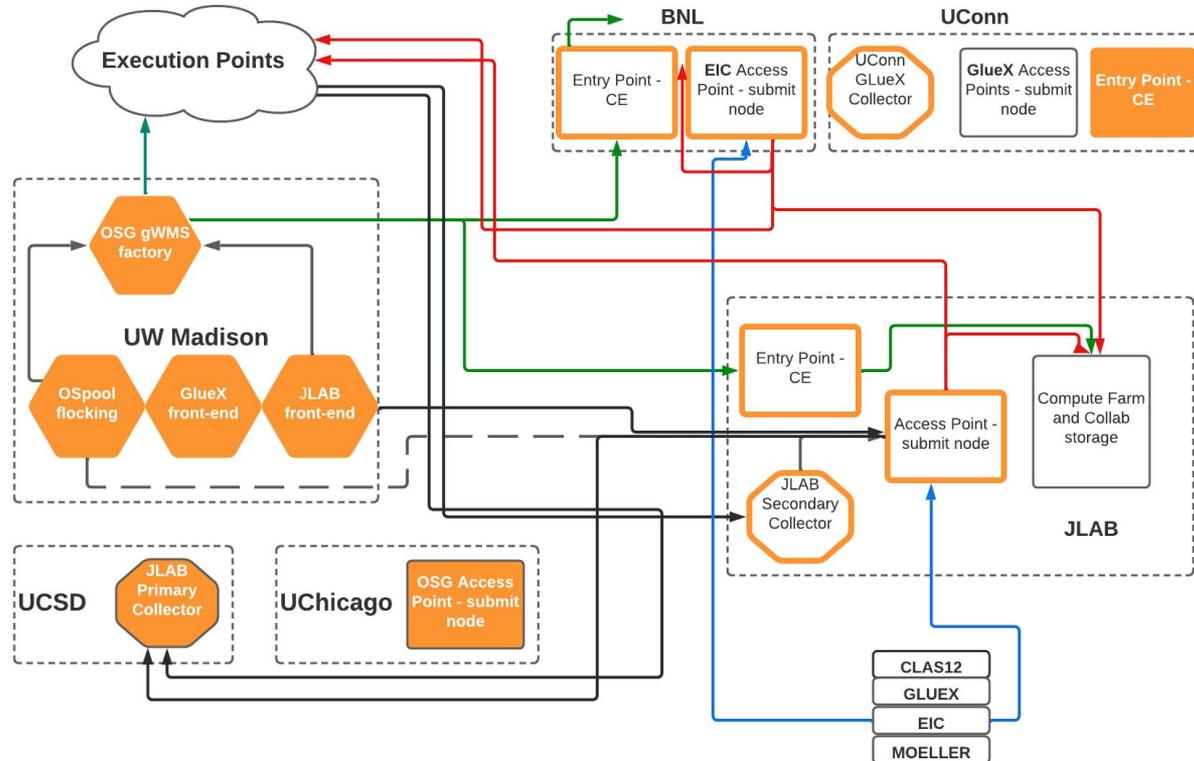
JLab Experiments on the OSG



JLab Experiments on the OSG



Example: JLab Experiments on OSG



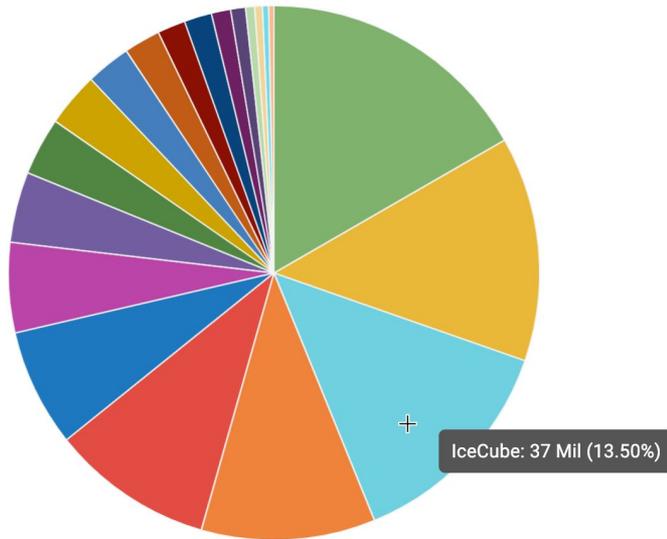
- **A single collaboration may require engagement effort to multiple facilities**
- **A single facility may require engagement effort to multiple collaborations**

Lessons learned

- Focus on needs of collaborations provides opportunities to learn and adapt for the OSG team
 - Common problems inspire solutions that have value to many collaborations
 - Cascade of best practices, know-how and technologies
- Large collaborations (e.g. ATLAS & CMS) paved the way.
 - Technologies developed there, e.g. Rucio, Xrootd and HTCondor-based distributed processing frameworks etc. are leveraged by OSG for less resourced collaborations
- Partnerships between collaborations-OSG-facilities help inform future projects
 - Collaborations can emerge and grow within the scope of future facility plans
 - Partnerships with OSG allow technology transitions and leverage between consecutive science collaborations ... e.g. SPT-3G => CMB-S4

Compute resources Utilization (past year)

Core Hours per collaboration



	total
dune	46 Mil
CLAS12	38 Mil
IceCube	37 Mil
microboone	29 Mil
nova	27 Mil
gm2	20 Mil
LIGO	15 Mil
REDTOP	12 Mil
minerva	10 Mil
spt.all	9 Mil
des	7 Mil
mu2e	6 Mil
GLUEX	5 Mil
icarus	5 Mil
EIC	3 Mil
ligo	3 Mil
sbnd	2 Mil
seaquest	1 Mil
gluex	1 Mil
xenon1t	896 K

- 270 million core hours
- Mix of Open Science Pool and collaboration specific resources
- About half on FermiGrid
- Small collaborations are treated like individual researchers while larger ones are dealt with by collaboration support team.

Examples of Collaboration Storage

- 330 TB used on PATh operated storage - Snowmass, REDTOP and XENON
- 2.1 PB of storage at UChicago for the XENON and SPT-3G collaborations
 - UChicago provides institutional support for Xenon and SPT-3G
- 6 PB of storage managed by Rucio for XENON across US (3) & EU (5) institutions
- 800 TB of data for SPT-3G distributed across 4 institutions
- 5 PB for LIGO observational data, reduced datasets and products at Caltech
- 2 PB of storage for the EIC collaboration at BNL and JLAB
- 7 PB available at JLAB for CLAS12 and GlueX
- 400 TB for VERITAS/CTA at Gatech

~25PB stored across a dozen institutions for 10 collaborations, stored in systems implemented via different technologies.

Summary & Conclusion

- OSG continues to **advance all of open science via the practice of dHTC, and the advancement of its state of the art.**
- Support across many science domains, large data and compute intensive workflows
- Open Science Pool as strategy to **democratize access to dHTC**
- Collaborations are an important component in the dHTC ecosystem of OSG. They add capacity to the distributed resources and contribute to better anticipate and prepare for future expectations and augmented scale in service deployment and management