



Year 6 Priorities

Frank Würthwein
OSG Executive Director
Professor of Physics
UCSD/SDSC



- We're still at the very beginning. So by no means do we have a fully fledged out set of priorities.
- What I have put together for these slides is just the snapshot of our present thinking.
- Feedback is more than welcome!

Top Priorities

- Integration
- simplify, simplify, simplify
- growth
- maintenance and operations that are just business as usual.
- networking as a strategic focus
- data capabilities

- We got work left to do in our AWS integration offering via condor_annex
 - Want to have deadline use case fully implemented and in production.
 - Have Vanguri as use case. Trying to figure out who else!
- Want to have more XRAS requests for Comet, Stampede, and Jetstream
 - this implies making sure that we have all necessary integration done.
 - xrootd cache in front of Comet for ATLAS, LIGO, and others?
- Backfill on JetStream
- What goals can we have for DOE supercomputers?

Year of the Retirements!

- Somewhat unexpectedly, many of our (software) friends have been retiring en masse:
 - **GRAM**: Already gone from most sites for about a year.
 - **glexec**: To be replaced by a new component.
 - **GIP/BDII**: Replacement (OSG Collector) already integrated into HTCCondor-CE. You haven't used this in awhile.
 - **Gratia (central-only)**: Move from a monolithic MySQL database at FNAL to a decentralized architecture. Database is ElasticSearch at Nebraska.
 - **bestman2**: Replaced by load-balanced GridFTP.
- Ideally, retiring components frees up your time to do other things!

- As a positive side effect of getting rid of glexec, we will have the singularity, and thus container technology as a first class citizen in OSG.
- There will undoubtedly be work in thinking through what this means
 - does it increase the amount of data transferred with jobs?
 - does it lead to scalability problems?
 - does it expose scratch space limits at sites?
 - ...many other things we haven't thought off, maybe.

Simplifying the VO Zoo

- Setting up a classic VO is hard: **Why would you do that?**
 - *Policy enforcement*: **sites** can enforce policies specific to a VO; **VOs** can directly manage their share of resources.
 - *Isolation*: you do not want other VOs to interfere with your payloads.
- Singularity is one mechanism to provide isolation without needing a separate VO.
- In general, policy enforcement is *difficult*. However, we have tools for many simple policies
 - Particularly, cases where site is “owned” by a single VO and everything else is opportunistic.
- The “support matrix” has (# VOs) * (# site) entries. Decreasing number of distinct VOs *as seen by the CE* saves effort overall. **Do you really need to submit your own pilots?**
 - Default GUMS template is **2,000 lines of XML** and **authorizes about 20,000 users** at the CE. We can do much better!

Looking forward to working with the community!

simplify, simplify, simplify

- Understand what's left to retire GUMS
- Get rid of all needs for x509 certificates in browsers.
- get rid of OSG twiki
- get rid of RSV
- get rid of GRATIA
- what else can we get rid of to reduce services ops maintains ?

- Focus on growing resource pool any way we can!
 - LIGO India ~ 2000 cores
 - glueX collaborators: 29 institutions, some of which are outside US
 - VERITAS collaborators: 20 institutions
 - XENON1T collaborators: 22 institutions
 - IceCube collaborators: 40+ institutions
 - SPT collaborators: 20+ institutions
 - LHC collaborators: 80+ institutions
 - capitalize on campus champions & CaRC communities
 - who else?
 - Are there obvious targets because they include multiple experiments at the institution?

Aside

- We believe that the primary vehicle for growth is the hosted CE.
- There is undoubtedly work related to this that we haven't thought through yet.
 - how many can we support?
 - how do we support them?
 - where are they hosted?
 - how do we pay for this hosting?
 - ...add your favorite concern

- Transition to SL 7
 - probably other things that I am not aware of and forgot to ask
- There is the usual ops that needs to get done.

- We took the network performance data store off production because the implementation is too fragile to meet the SLA.
 - poor software choices that we went along with because ESNNet chose them and we wanted to be a good collaborator.
- Considering making network performance measurements within HTCondor available to public -> High on NSF's wish list.

data capabilities

- we benchmarked HTCondor file transfer last year as our primary solution for output data handling
- we introduced stashCache last year as new offering for more efficient input data use
 - this will require continued effort to roll out, understand use, train, monitor use, etc. etc.



Open Science Grid

What else?

