OSG Technology Area

Brian Bockelman

OSG Staff Retreat 2018

Technology Investigations?

- In the OSG N5Y grant, we had a dedicated Technology Investigations team to guide the technological evolution of the OSG.
    - The team has a great track record over the course of the grant.
- What role does Technology Investigations play in the OSG-Core model?
    - This is an area that strongly overlaps with the R&D components of participating projects, meaning the ability to drive the work is significantly different than in the past.
        - **Example**: If GRAM had lasted until 2018, could we have invented the HTCondor-CE in 2019?
    - **Proposal/Discussion**: should this be the "*Technology Evolutions*" area to reflect that it is often coordinating the evolution of OSG as opposed to doing the investigations in the core?
- **Alternate observation**: We have more significant effort in data & storage management than we have for many years. Should we have a more explicit "storage team" (not area)?
    - We have a dedicated XCache vertical involving development & ops. Other places are horizontal.

Projects

- What projects — not already discussed — does the Technology Area need to work on over the next year?
    - Jupyter-based interface to the OSG.
    - Rationalize approach to XCache and StashCache.
    - Incorporate
    - R&D necessary to replace Globus Toolkit:
        - Help mature GridFTP replacements (HTTPS, Xrootd).
        - Piece together an overarching strategy to replacing GSI/VOMS.
- Important projects discussed already: GRACC “scrubbing”,

Jupyter Interface

- Jupyter provides an alternate UI to computing infrastructure. “Notebooks” versus “terminal”.
    - Some indication there’s demand - or upcoming demand - for this interface to be provided by OSG.
        - In the terminal UI, there’s a clear delineation between when interactive ends and batch begins. Jupyter … it’s not as clear.
- What does OSG want to do here? Can we determine this here or at a Blueprint?
    - Do we have enough expertise to do a Blueprint on this?

Rucio, XCache, StashCache

- Input I need for planning the next year (unclear how much we'll need this slide after preceding storage discussion):
    - **Rucio**: What's our goal? Do we want to improve our support of the existing, identified, mid-scale users (LIGO, IceCube, etc) or
        - **Example**: there's a set of work items
            - Observation: all Rucio-related items were left out of Operations
    - **XCache**: What's our best way to provide value? We have a XCache meeting but it's not clear it is valuable to anyone outside StashCache…
    - **StashCache**: Clear work items needed around automation and preparedness for including more origins (particularly for authenticated instances).
        - Recent CHTC participation has been key!
        - How far do we want to drive the containerization-based deploys here?
            - **Note**: we don't have a released version of XCache stable under heavy load. Makes it very difficult to do "traditional" deployment models.

GT replacement: Transfer Protocols

- In January 2018 blueprint, we agreed on:
    - Focus on HTTPS/WebDAV as the OSG approach for third-party-copy.
        - Xrootd protocol for streaming
- 9 months later, do we want to revisit this?
- In the meantime, initial technical capabilities have made it to the OSG Software.
    - Next step is documentation and integration tests.

GT replacement: GSI/VOMS

- We build our auth{z,n} system out of three layers:
    - **X509/TLS**: Authenticate a TLS session by using client certificates.
    - **GSI**: Provides concepts of delegation and augments “vanilla” PKI used for TLS (namespacing of CAs, various traditions like CRLs, etc).
    - **VOMS**: Provides group / attribute information.
    - We then take the resulting authenticated global identity/group information and map it to a local identity (Unix user) and authorize it as that local identity.
- SciTokens has demonstrated the ability to switch to a capability-based system for data transfers. WLCG Authz working group effectively endorsing this approach (with minor tweaks to the format).
    - For next year, goal is to land / expand these capabilities in OSG for storage.
- How do we move forward with the rest of the ecosystem? How do we get away from GSI/VOMS at the CEs? Should we

What projects are left?

- If we take the list from our brainstorming document, here are the items not discussed at time of writing:
    - Add Open Storage Network (OSN) storage as a StashCache Origin.
    - FIONAs for opportunistic computing.
    - Slurm-Glidein Integration.
    - Scale test NDN.
    - Support for interactive Machine Learning on OSG.
    - Eliminate the need for CVMFS on worker nodes.
- Which should we pick up and work into the plans?

Blueprints

- What Blueprint meetings do we need in the next year? My suggestions:
  - Post-Globus auth{z,n} strategy (focusing on CEs).
  - Strategy on providing a Jupyter service.
- *Possibly not a blueprint*, but a coordination meeting on organizing XCache activities?