

OSG Space Management

Introduction

This document describes a pivot from “opportunistic usage” of storage elements on the OSG’s storage elements. The opportunistic usage paradigm was that VOs would request an amount of space for a desired period of time; this would be done through SRMv2 space tokens. After the space expired, the VO’s files would be automatically cleaned up unless the VO requests an extension.

This document outlines a new, less-powerful paradigm simply termed “space management”. This idea focuses on simplicity,

Issues with Opportunistic Usage

We outline issues we perceived with the current opportunistic usage model:

1. **Lack of support from SE vendors.** In the OSG, only dCache has a strongly integrated SRMv2 component. However, current trends are going toward best-of-breed systems developed outside the grid niche (Xrootd, HDFS, cluster file systems such as Lustre or GPFS); these rely on modular SRM components such as BeStMan-Gateway that does not support true space reservations. Further, the usage of opportunistic storage (especially areas being automatically cleaned up) even in dCache has been rare as the WLCG only specifies the need for static space tokens.
2. **Interactions required between sites and VOs.** Due to the complexity of opportunistic usage, space usually needs to be requested per-site-per-VO. Each VO would need to contact each site individually over email to negotiate and setup usage. Even in cases where usage was negotiated en masse (i.e., DZero usage of CMS sites), setting up of areas was done one-by-one.
3. **Overly-complex to setup and use.** For VOs, there was no single set of tools that could be used to manage spaces that worked at every site. The learning curve was too high and the entire system is too fragile. The biggest success, D0, required dedicated coordination with the OSG VO Group, significant development to adopt the DZero SAM system to use SRM, and full-time personnel dedicated to maintaining the setup. With this, DZero uses 5 OSG SEs; 2 of these have no support for space tokens.

We state these three issues up front because we want to avoid them in the next round of SE usage. Specifically, we want a system that works regardless of SE vendor support, does not require interactions between sites and VOs, and is relatively simple to use.

Space Management and Public Spaces

We introduce two new concepts, “space management” and “public spaces”:

- **“Space management” is the ability to monitor and manage the use of space at a site’s storage element.** This includes a wide variety of techniques, including SRMv2 space tokens, traditional Unix quotas, and the usage of SE-independent storage accounting probes. The important pieces of space management are:
 - **Site-centric.** The focus is to make sure sites understand the usage of their space. These tools are not aimed for consumption – or usage – by VOs.
 - **Widely deployable.** All sites should be able to manage the space at their site in some manner – regardless of what’s present
- **“Public space” is a small amount of space set aside at a site that non-owner VOs can use.** We target this to be approximately 5% of the size of the site. This area should be treated in a similar manner to \$OSG_DATA on the CE. No pre-existing arrangements should be needed for usage, but if the space is exhausted, it is gone until it is manually cleaned up. There is no QoS guarantee for this area, but sites should not automatically clean up old files. If a file deletion must be done, the site should notify the VO (preferably, but not necessarily done before the file deletion) and delete all of the VO’s data. A site should not delete a subset of the data – often, re-transferring all files is much easier than determining which ones are missing. Important attributes of public space are:
 - **Any VO can use it without prior arrangements.** The top-level directory should be world-writeable and each VO should use a sub-directory with the VO’s name. It is important to prevent sites from adding barriers that require individual VOs to contact them.
 - **No per-VO limit is necessary.** Per-VO limits may be implemented, but are not necessary – a site can treat this as OSG community space.
 - **Correctly advertised in the OSG information systems.** The space should be discoverable through the BDII and ReSS. The most important attributes are (in order): SRM endpoint and directory, whether the SE is POSIX-mounted on worker nodes, and the amount of space remaining.

Where necessary and convenient, the OSG will negotiate the construction of public spaces from resource providers. Focus will initially be given to sites that offer POSIX-like access.

Interaction with the OSG

The OSG will take on two important responsibilities:

1. Facilitate in the creation of public spaces (OSG-ET). The OSG-ET will work with the larger resource providers (USCMS and USATLAS) to determine what sites will provide public spaces and how large these spaces will be.
2. Assist VOs in understanding and using the technologies (OSG-Storage and Engagement). OSG-Storage will develop a survey to help the OSG understand how VOs would like to use space; this will help us understand whether they prefer a “hands-on” or “hands-off” approach to assistance and help avoid usage patterns that do not fit well with the OSG. For example, a VO that expects it will move millions of small files to many sites with minimal effort will fail regardless of the amount of effort OSG or individual sites contribute.

By assisting sites and VOs where necessary, the number of lines of communication for N sites and M VOs is reduced to N+M instead of N*M.

Recognized Problems

Even with the proposed solution, there are a few significant issues facing OSG-Storage that remain unsolved:

1. **Introduction of new VOs.** This is an issue facing the OSG as a whole; the lead time for the setup of a new VO and getting authentication at sites can be staggering. Even once the VO has been set up, rolling it out to individual sites can be time-consuming and labor intensive on the VO's part.

For sites to add a new VO, they first must be contacted by the GOC or the VO and asked to take action (GOC regularly announces when new VOs are added, but these are just general announcements and generally not followed). This is the most time consuming part as each site needs a contact. The site has to edit the GUMS configuration (which can have serious negative side-effects for site stability, especially in older versions, causing sites to be reluctant), set up user accounts, and home directories. Each step can possibly fail, and generally must be restarted for each site.

2. **Transfer management.** The OSG has no generic mechanism to transfer large datasets from one site to another. Each VO that has automated transfer mechanisms have developed their own software independently with little-to-no overlap. New VOs have no off-the-shelf software available to them that works with the OSG. The closest available software is Globus's RFT (doesn't support SRM), Stork (doesn't support SRM, but under active development), and FTS (very complex server that requires Oracle; historically changes require long development times).
3. **Adopting applications to grid SEs.** Not all grid SEs support POSIX access. This means VOs must have flexible software layers (uncommon for all but the largest organizations) or limit themselves to a smaller subset of OSG SEs.

Roll-out of Space Management program

1. Immediate usage of “seed sites”. Several sites are ready for immediate participation in this program. This should give us immediate feedback to determine issues we may have overlooked.
2. Have OSG-ET work to setup public spaces from the major resource providers.
3. Develop space management tools for the major SEs. This would involve the OSG Storage, Metrics, and Documentation teams. Some of the tools are pre-existing and just need better documentation (Hadoop quotas, dCache space tokens), while others (POSIX and dCache Gratia storage probes) need to be developed. Ideally, each SE would a well-documented and supported tool for space management.
4. Have a mechanism to test the public spaces at sites. This will not be continuous testing like RSV, but a targeted tool for verifying that site’s public spaces work.
5. Update the OSG documentation to reflect the new space management paradigm.
6. Work with the VO group to make sure any new VO fills out the VO questionnaire if they would like OSG-Storage assistance.

VO Questionnaire

1. Part of Engage or a separate VO?
2. How many files to store, how many files per directory?
3. For how long do you need to store these files?
4. What is a typical size of the file?
5. Does job need to access file locally?
6. What is time duration of the job?
7. What is min access throughput?
8. How many requests and of what kind to the storage per job?
9. If you need to copy files to the local disk what is size of the “local area”