

STORAGE OF OUTPUT FROM JOBS THAT REMOTELY

Doug Benjamin
Duke University

Analysis I/O

- Most users will start from centrally produced derived xAOD's
- The US ATLAS will attempt to place all “important” all derived xAOD's in the US
 - This was attempted with Group D3PD's in Run 1
 - Of course this is disk space dependent
- US ATLAS analyzers will have 5-10 TB of disk space spread across LOCALGROUP disk located in the T1 and T2 sites.
 - This requires Rucio to handle group quota's to be manageable

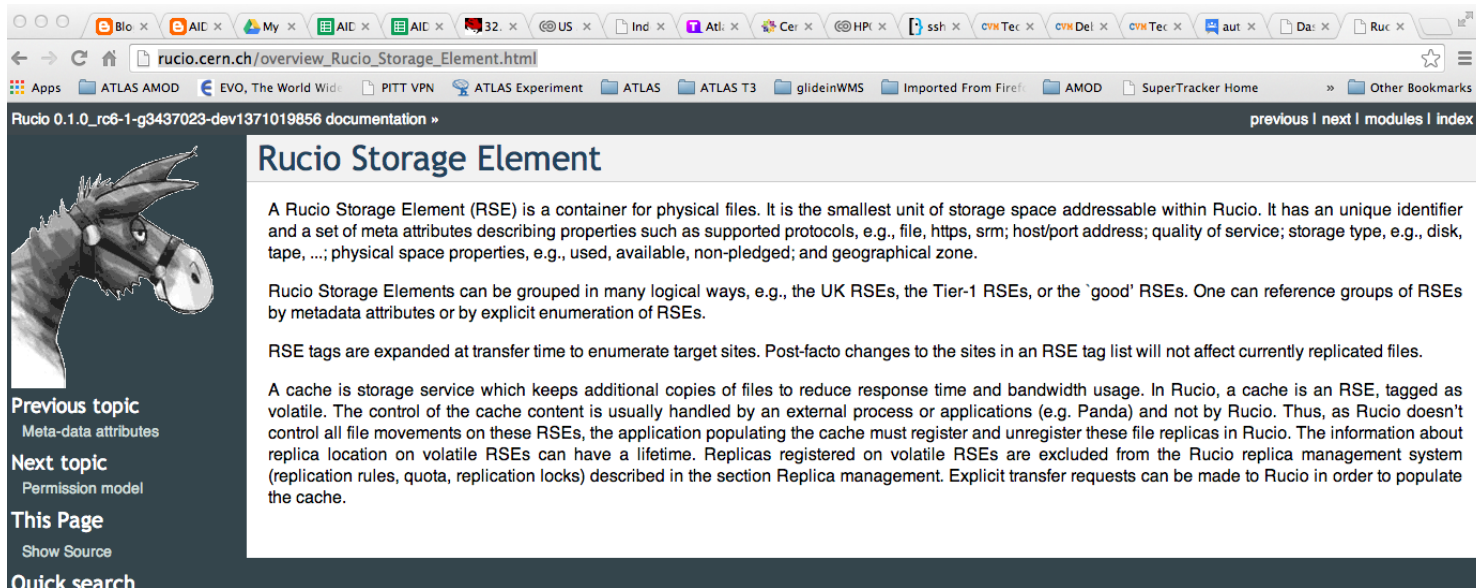
Analysis I/O (continued)

- Some US ATLAS groups have sufficient local disk space to have their own LocalGroupDisk (UPENN, LBL for example)
 - These users want to and have the ability to direct their PanDA output directly to their local disk resources.
 - ATLAS tools are supposed to be used to manage the storage
- Many US groups have 10's TB disk space and not 100's of TB
 - Some of the users in these groups want to automatically copy PanDA output to local resources
 - These groups have a small fraction of a local person for storage system support

Rucio Storage Elements (RSE)

- Large groups continue to use the standard Rucio Storage element (RSE) and existing technologies (SRM and gridftp server). LOCALGROUPDISK
- Smaller groups (most of US ATLAS) need lightweight Rucio Storage Element
 - Requirements – easy to maintain
 - Has a minimal effect on Central Computing operations
- Client tools used for the fetching smallest datasets
- Remote data access for local computers (cluster or desktops – even laptops) using FAX (federated storage) is a game changer
 - Will this reduce the amount of data transferred to local resources?

Rucio Cache



Rucio 0.1.0_rc6-1-g3437023-dev1371019856 documentation »

Rucio Storage Element

A Rucio Storage Element (RSE) is a container for physical files. It is the smallest unit of storage space addressable within Rucio. It has a unique identifier and a set of meta attributes describing properties such as supported protocols, e.g., file, https, srm; host/port address; quality of service; storage type, e.g., disk, tape, ...; physical space properties, e.g., used, available, non-pledged; and geographical zone.

Rucio Storage Elements can be grouped in many logical ways, e.g., the UK RSEs, the Tier-1 RSEs, or the 'good' RSEs. One can reference groups of RSEs by metadata attributes or by explicit enumeration of RSEs.

RSE tags are expanded at transfer time to enumerate target sites. Post-facto changes to the sites in an RSE tag list will not affect currently replicated files.

A cache is storage service which keeps additional copies of files to reduce response time and bandwidth usage. In Rucio, a cache is an RSE, tagged as volatile. The control of the cache content is usually handled by an external process or applications (e.g. Panda) and not by Rucio. Thus, as Rucio doesn't control all file movements on these RSEs, the application populating the cache must register and unregister these file replicas in Rucio. The information about replica location on volatile RSEs can have a lifetime. Replicas registered on volatile RSEs are excluded from the Rucio replica management system (replication rules, quota, replication locks) described in the section Replica management. Explicit transfer requests can be made to Rucio in order to populate the cache.

Previous topic
Meta-data attributes

Next topic
Permission model

This Page
Show Source

Quick search

- **Lightweight storage element**
 - Storage system information (ie files) sent to CERN and captured by activeMQ message bus
 - Still in development
 - Once Rucio is working, more effort will be available to finish development, integration, testing and deployment
 - Gives the smaller US groups the functionality they request
- Caching space is also useful at larger ATLAS sites (T1, T2's) - not a bad idea in my opinion

Yet another 3 letter acronym

- The current Tier 3 White paper talks about LRU (local Resource Unit)
 - Amount of Local CPU's and disk space that an analyzer needs to efficiently analyze derived xAOD's at her institution.
- Is this space ~ 10 TB going to be PanDA aware (ie will PanDA be able to schedule against this storage trivially)?

Questions for discussion

- Do the smaller US ATLAS institutions need its local storage to be DDM compliant?
- How do we (ATLAS) provide a solution for the smaller groups that provides what people need and is ready preferable at the start of Run 2 or early in Run 2?
- If groups don't want DDM compliant local storage, do we have the tools that allow users to process datasets at their local storage?
- Now do we access local storage (Read/Write) when “flocking away” to beyond pledged resources in the US ATLAS Tier 1 and Tier 2 sites?