

Chapter 6:

Experiment perspective

Content

- This talk will present the experience gained in operating Rucio in ATLAS
- What will be presented
 - Synchronization with other services
 - Integration of Rucio with Workflow Management system
 - Naming convention (scope, PFN)
 - Migration to Rucio

Synchronization with other services

- Rucio can be interfaced with different services to handle
 - Creation of new sites and management of protocols
 - Creation of new accounts
 - Quota and permission management
 - Space reporting
- All these synchronizations are run as scripts i.e. probes on a nagios node
 - Automatic notification in case of problems
 - Retries

Synchronization with identity service (VOMS)

- Accounts creation automatized
 - A probe collects the list of identities in VOMS and the list of accounts registered in ATLAS group at CERN
 - It creates the account in Rucio and associates to this account the VOMS identity (based on nickname)
 - It creates a scope user.<username>
- Account privileges
 - Privileges are based on roles in VOMS (e.g. /atlas/role=Production → admin in Rucio)
- Country and physics group ownership
 - Collect the different groups in VOMS and their associated identities and map them to GROUP accounts in Rucio
- Automatic quota scaling with space available at the site based on the space reported by each site
 - Specific probe to collect the space deployed at the sites

Synchronization with information system (AGIS)

- AGIS is the information system of ATLAS that stores
 - List of all ATLAS sites
 - Ordered list of all the protocols supported by these sites
 - “Closeness” between the sites
 - Different metadata like country, DISK/TAPE, quota for physics groups, etc.
- AGIS is also responsible of collecting the site downtimes
- Automatic collection of all these information into Rucio

Synchronization with the Storage Elements

- Collection of the space used/free via storage provided value or json file published by the site
 - To define quota
 - To identify possible issue with the Storage Element (consistency metrics)
- Automatic consistency check between the files recorded in Rucio and the files stored on the Storage Elements (SE)
 - Sites are asked to publish regularly Storage Element dumps
 - Dumps are automatically fetched by Rucio and compared to the list of files recorded in Rucio
 - 2 types of inconsistencies can be found :
 - Dark Files (i.e. on SE but not known by Rucio) are cleaned by a special Rucio daemon
 - Lost Files (i.e. in Rucio but not on the SE) are declared to the Rucio recovery service

Synchronization with metadata service

- Rucio is interfaced with ATLAS Metadata Interface (AMI) to report
 - File loss
 - New dataset attached to container
 - Change of DID state : open → close
- The synchronization Rucio → AMI is done using ActiveMQ
- On the other hand, Rucio gets from AMI
 - New scopes created
 - Datasets and containers naming convention

Integration with Workload Management

- Rucio is integrated with ATLAS Workload Management System (PanDA/Prodsys)
- Rucio takes care of
 - Transferring of data (input/output) for the jobs
 - Staging data from TAPE
 - Notifying Panda when transfer of dataset is done
 - Notifying prodsys in case of lost or corrupted files
- Rucio also reads the PanDA tables to identify the datasets currently used by the jobs and can make extra copies to improve the brokering
- Rucio can possibly be used by any other Workflow Management system

PFN convention

- ATLAS convention
 - LFN : mc15_13TeV:EVNT.13080669._137329.pool.root.1
 - PFN : mc15_13TeV/66/a2/EVNT.13080669._137329.pool.root.1
 - 66a2 : first 2 bytes of md5 hash of mc15_13TeV:EVNT.13080669._137329.pool.root.1
- Question : Why did you choose this convention?
- Response
 - Deterministic : One just needs to know the LFN to locate the PFN
 - Hash part
 - Constraint that some storage technologies can not handle more than 32k files in a directory
 - Hash allow to have a well balanced distribution of files in directory
 - Scope part
 - To make the space more human readable
 - For permission on the storage element

PFN convention

- Pros and Cons :
 - Pros :
 - See previous slide
 - Cons :
 - Too many (empty) directories. Up to $65k * nb_scopes$
 - Cryptic for users
- Conclusion : Naming convention is something important that deserves careful thinking (very difficult to change after)

Migration experience (1)

- Rucio has been fully put in production in December 2014
- At that time, there were already ~300M files registered in the previous Data Management service of ATLAS (DQ2) on ~100 sites
- Strong constraints to perform the migration with limited downtime
- Migration strategy thought well in advance to move gradually to Rucio. 3 steps :
 - Moving the files at the sites to follow the new deterministic path convention
 - Migrating the file replica catalog from LFC to Rucio
 - Migrating all the dataset, containers, subscriptions/rules from DQ2 to Rucio

Migration experience (2)

- First step of the migration ran in background mode for ~1 year
 - File moved (metadata operation) centrally by a dedicated infrastructure
 - Include file migration on the site and in the LCG File Catalog (LFC)
 - 300M/1 year ~ 10 Hz
 - No downtime
- Migration of the file replicas from the LFC to Rucio
 - Production/analysis stopped for one day cloud by cloud (i.e. only ~10 sites not available at a time)
 - PL/SQL procedure
 - Deployment of a DQ2 client able to talk to Rucio servers

Migration experience (3)

- Final step : Migration of the content of DQ2 to Rucio
 - One full day downtime for all the ATLAS sites
 - PL/SQL procedure
- Lessons learned :
 - Migration from one Data Management system to another one is possible
 - without long downtime
 - even for a big collaboration
 - But needs a lot of careful planning

Questions?

If you have a question but don't get the chance to ask it directly during the session, you can do it here: <https://goo.gl/BdSGoC>