CMS/CERN IT storage meeting

FTS

Steven Murray - mentioned Filipe's meeting with some of the CERN FTS team. Felipe has observed a drop in transfer performance when he submits multiple containers/datasets to RUCIO in parallel when compared to submitting them one at time, with one being allowed to finish before the next is submitted. The FTS team reported that FTS CMS at CERN has not had any performance problems concerning the volume of CMS transfers. The performance problems associated with submitting multiple containers/datasets to RUCIO in parallel should be looked at by both the CMS Jira developers and FTS.

Katy - the long gap between FTS and Rucio completions - could be fixed by using a different component in Rucio (according to the Rucio dev meeting last week) but to use that we need FTS to indicate via TransferComplete messages which transfers have archive monitoring enabled (this work is scheduled).

Staging errors - FTS has reported some file transfers as having failed because files successfully recalled from tape have been garbage collected before they could be copied to another storage end point. Examples of this problem have been seen when transferring files from Antares to Echo at RAL. Maybe RAL should look at how long garbage collection takes? Is data going to Echo being slowed down?

CTA

Staging *leak*

Similar types of staging failures are taking place at CERN:
- A file is staged from tape in the tape buffer
- By the time the transfer job tries to read this file out of the tape buffer it has been removed

Investigations need to be conducted at CERN:
- Current measured leak is around 24% at 3GB/s staging rate

CTA could configure some disks from CMS pledge to absorb this staging leak:
- Julien to define the amount of disk needed for this task
- Dima will get the green light from James and Danilo

Monitoring this *leak* will allow us to better understand it (multihop, destination site,...) as this problem was already identified during the previous B-parking recall and not investigated further once this activity was over.

Massive timeout staging failures

CMS is using the default 7 days *bring_online_timeout* value for massive recalls that will need much more time to be recalled.

Julien suggested increasing bring-online time out from 7 days to 30 days.

Recalling files outside of parking staging activity

Different (Rucio) activities for different CTA-recall priorities - can Workload Management use this? E.g. to allow other activities to get some bandwidth when there is a Parking recall.

Two approaches are possible:
- CMS DM should tell users to create Rucio rules in a specific activity different from parking activity
  - This higher priority activity should be communicated to tape sites so that they can map it to higher priority staging queues
- Use a dedicated priority for parking staging and communicate it to tape sites so that they can map it to lower priority staging queues

Archive metadata discussions starting

CTA is helping CMS with tape families for legacy data but it is not good enough to tackle the parking data grouping. The new direction will be to flag data with archive metadata to indicate tape sites how to better group data (HTTP transfers).

At least 2 possible areas of interest from CMS:
- Grouping parking data
- Lower *archive_priority* for MC data

Broken tapes on CTA

Julien will set the broken files to 0 size and will start publishing weekly namespace dumps identifying broken files. As a consequence of this inconsistency the flagged files won't be revalidated in CMS Rucio NS anymore.

As a next step he will contact Igor to invalidate these files from T0 tape.

Finally the CMS DM will decide how to handle these files: to overwrite/delete and sort it on the Rucio side.

Additional questions:

Is 'archivetest' still mapped to an active RSE in CMS?

Space in EOS - Steve Murray is talking to Luca who says CMS is only using half of its space…this is mysterious - CMS to investigate why the free space is not being used?