



ALICE MSS policy – implementation and plans

Latchezar Betev (ALICE)
GDB, February 6, 2008

File archives

- Archive option for files:
 - OutputArchive
{AliESDs.root,AliESDsFriends.root:root.zip@somestorage}
 - reserved for ROOT files (zip -n)
 - OutputArchive{*.log,stdout,stderr:logs@someotherstorage}
 - zipped on the fly – all other files
- In production since 2 years
 - Fewer (and larger) files in storage – especially important for MSS
 - Applicable for all storages
 - No loss of performance for ROOT files – direct access of archive member

File archives (2)

- Production files – average gain 6/1, average file size in MSS from MC production is 200MB
- User files – not controlled, users are encouraged to group all output files in archives

Post-processing

- Aggregation of ESDs from multiple jobs
 - One run typically consists of few hundred RAW data chunks, processed as independent jobs
 - Each job ESD is put on a temporary storage (disk) and aggregated at the end of the run processing
 - Aggregated ESD is put in MSS, intermediate files deleted
- *Implemented, but not put in operation*
- *Major issue is the management of the temporary storage*

SE registration threshold

- Small files in CERN CASTOR2 – written, but never read back
 - Many thanks to the CASTOR experts for pointing this out!
- Problem – failed jobs registering empty archives (nobody ever read these out)
- SE threshold implemented – storage refuses to store a file below threshold size
 - Applicable only for MSS
 - In production

SE registration threshold (2)

- Problem – what to do with small files
- If smaller than registration threshold, automatically redirected to disk based storage
 - What if it fails?
 - Simple replication does not help – users require custodial storage for important data
- Effective MSS threshold cannot be put very high
 - Unless there is a complex aggregation mechanism for all files stored

File size: present status and plans

- RAW data – largest volume of data in MSS
 - Current size of one chunk – 1GB
 - CCRC'08 – 2GB
 - May – 10GB
 - With this size, the 'post processing' is not needed, ESD size 1GB (10% of RAW)
- ESDs – second largest
 - MC production average size – 200MB
 - Size is function of #events in one job, can go to x3
 - RAW data – 100MB (with 1 GB RAW)
 - Post processing or larger RAW data chunk size
- User files - ?

Access to data in MSS

- RAW data – exclusively by production jobs
 - Pre-staged in large chunks
 - Not considered an issue
- ESDs/AODs for user analysis (if data only in MSS)
 - Pre-staged datasets for high statistics processing
 - Large scale analysis of non-staged data is strongly discouraged
 - Users make (honest) mistakes and do from time to time a large data staging on their own

Summary

- Use of file archives for all storages – in production
- SE threshold for MSS – in production, need to find effective threshold value
- Post-processing and aggregation – implemented, but difficult to manage on Grid storage
- Analysis – managed datasets
- File sized – increase size of RAW to 10GB, ESD to 1GB (May CCRC)