Laboratoire d'Annecy-le-Vieux de Physique des Particules

Implementing WLCG diskless sites for production

S. Jézéquel

28 May 2017

ATLAS EXPERIMENT

cnrs

UNIVERSITÉ SAVOIE MONT BLANC

Introduction

'ATLAS recommends sites with small storage (< 400 TB) to focus hardware investment in CPUs instead of storage'

→ Diskless sites (excluding LOCALGROUPDISK)

- Policy proposed by Eric Lancon when was Computing Coordinator
- Motivation: many small storages for production/analysis
  - → requiring similar human support at site and ADC ops as 'big' sites
  - → Not optimal usage of manpower and hardware (low site efficiency for prod)
- Input/output files accessed from remote SE ('diskless site')
  - Similar approach as 'cloud' sites
  - Another option : ARC-CE caching mechanism
- Global policy endorsed at last ICB (17 April)
  - But technical implementation was requested

Time to implement this policy (or give up)

Proposed implementation

- **Migration to diskless has to be endorsed by ICB representative (in contact with site)**
  - Critical to understand the impact on funding (possible reallocation) and HR
- Requirements for remote SE (rely on local knowledge) :
  - much bigger to sustain additional load
  - Stable+minimal network connection
- Simplest approach: minimise network usage and load on SE

→ Default configuration :

- No ANALYSIS as remote access is not available for analysis.
- PRODUCTION restricted to evgen/simu
- *More ambitious setup requires technical validation : endorsed by local teams*

Some details for local SE

- Link for size :
  - http://adc-ddm-mon.cern.ch/ddmusr01/T2_SCRATCHDISK.html
  - http://adc-ddm-mon.cern.ch/ddmusr01/T2_DATADISK.html

- Remains the support Grid storage for data transfers for local analysis batch

  (benefit from Rucio asynchronous file transfers) → LOCALGROUPDISK/SCRATCHDISK

- If local SE only used by ATLAS and no LOCALGROUPDISK

  → stop declaring SE in GOCDB before decommissioning the SE

Procedure and documentation

- First version of 'migration to diskless site' procedure documented in Twiki SitesSetupAndConfiguration#WLCG_site_migration_to_diskless
  - Based on M. Vamvakopoulos experience with 2 romanian sites (documentation)
  - Liaison between ADC/cloud support + local admins through JIRA follow-up ticket

- Technical steps
  - Check that WN can read/write files from remote SE (possible port number issue)
  - Update AGIS for Panda queues to stop ANALYSIS and use remote SE for PROD

    → No more ATLAS SAM tests on the SE
  - export primary datasets (mostly log) to remote SE and stop importing primary
  - Run over few weeks to validate setup
  - If OK, clean local SE with Rucio deletion commands

Small isolated T2 sites

- No action yet on isolated sites (DATADISK + SCRATCHDISK):
  - BEIJING-LCG2 : 310 + 60 TB : close to the 400 TB limit
  - TR-10-ULAKBIM (Turkey) : 131+5 TB
  - NCG-INGRID-PT (Portugal) : 177+10 TB : Could be technically linked to spanish SE
- Similar technical issue with many T3s (South Africa, Greece,...)
- Possible options :
  - Continue as today
  - ARC-CE caching mecanism

Very small T2 sites

- Started discussion with SE < 100 TB (my classification)
  - List reviewed at last ICB (8 sites = 10% T2 sites and 0.5% of T2 storages)
- 5 sites already agreed for migration to diskless :
  - Romania : RO-16-UAIC (fully done prod only), RO-14-ITIM (almost finished)
  - Russia : FIAN (has been broken for long period), ITEP (next test site)(9TB)
  - Austria : HEPHY-UIBK (OK since aiming to become 'cloud' site) (88 TB)
- Remains :
  - LUCILLE (US) (95 TB)
  - PSNC (PL) (44 TB): According to ICB rep, might become big in 2017 (~ CYFRONET)
  - SE-SNIC-T2 (Sweden) (30 TB) : Unused storage
- Similar review to be done for T3s (US, Germany,...)
- No significant impact on ATLAS if storage is decommissioned immediatly
- No LOCALGROUPDISK → No more Grid storage after SE decomissioning

Not very small T2 sites

- 100 TB < SE < 400 TB (10 sites : 12% T2s for 3% T2 storage)
- Russian federation :
  - RU-PNPI (169 TB) : ICB rep already suggested to become diskless
  - RRC-KI-T2 (334 TB) : Same support as T1
- Sites which will be reach 400 TB soon : RO-02-NIPNE(RO), FMPHI-UNIB (SK), WEIZMANN-LCG2 (IL)

- Remaining small sites per country (DATADISK + SCRATCHDISK):
  - USA : UTA_SWT2 : 120 TB , OU_OCHEP_SWT2 : 276+5 TB
  - UK : BHAM (225 + 20), CAM (235 + 20) + SHEF (340 + 13) : Internal UK discussion

  → Total : 1.2 PB

Conclusion

- **Concrete proposition for diskless policy implementation**
  - Clarification of list of 'small' sites almost done
  - Policy already implemented for 'very small' sites and will go on
  - Technical discussion still opened for isolated small T2 sites
  - Will be reviewed at next ICB

    →Mandatory to get feedback on daily work from ADC Ops or squad support

- T3s (not linked to T1/2) should be also reviewed

  for further ADC/site manpower optimisation
  - Should focus first on non-working sites