

Xcache activities in ATLAS

S. Jézéquel

Pre GDB - 8 July 2019

- * Critical component to go to current HL-LHC Strawman model
 - Also interesting for short term Grid organisation
- * Many infos extracted from last Ilija presentation : [DOMA ACCESS 18 June](#)
- * ATLAS documentation: <https://twiki.cern.ch/twiki/bin/view/AtlasComputing/XcacheATLAS>

* Interest in xcache :

- Grid production jobs
 - Caching : A priori negligible
 - Read-ahead : Smooth remote data access → Important to run high IO jobs (reco, derivation) → Replacement for small Grid SE
 - Reading full files most of the time
- Grid analysis jobs
 - Caching (of partial file content) : Under evaluation
 - Read-ahead : Smooth data access
 - Concept of hospital queues : Using controlled local queue to process data in replacement of remote CPUs
- Local analysis farm : Caching + read-ahead

- * Deployed in AGLT2, MWT2 and BNL
 - Slate AGLT2, MWT2, (NET2, SWT2 soon)
 - Still limited numbers of jobs (to be understood)
 - Local installation for BNL (SLAC soon) : Serve analysis facility
 - ESNET@Sunnyvale (soon for ATLAS users in West coast)

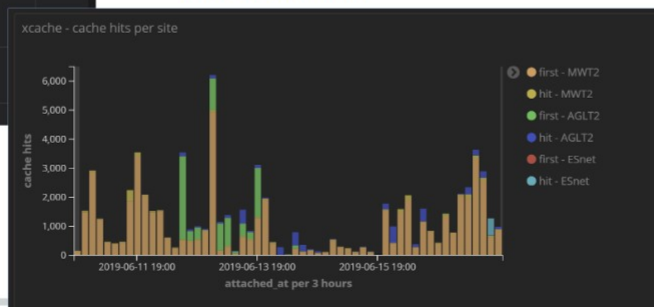
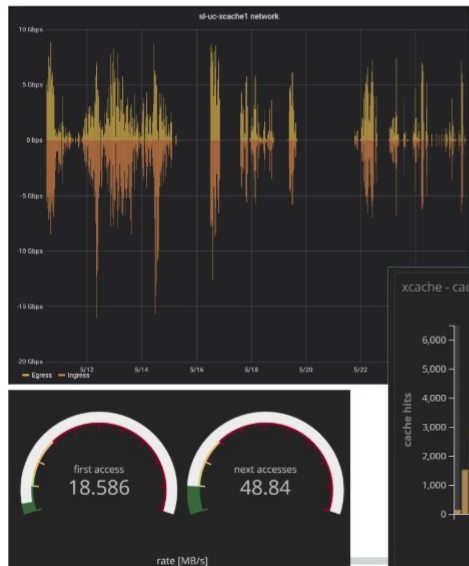
Performance

Stable. Discounting for SWT2, bad transfer checksums at permill level. That is still more than what we want.

Since this is a cache-unaware scheduling, cache hit rate is almost negligible.

Unexpectedly, load is quite spiky.

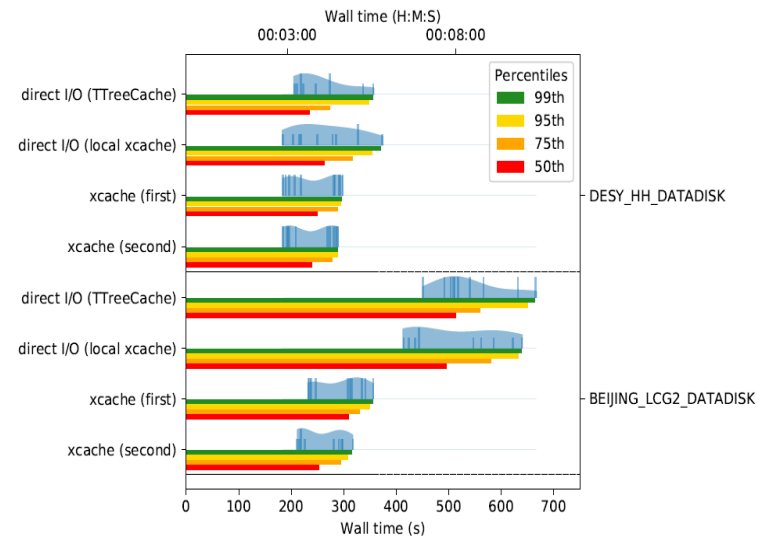
With 2 WQ threads and 10 WQ blocks/thread, cached file sparseness is at 1% level. Could be very different in realistic cache operation where we expect ~ 2 times more reads than writes to the disk.



- * Germany (LRZ-Munich)
 - Tested slate installation
 - Enabled to install xcache in few hours to keep Grid farm running while local Grid SE in downtime
 - Local installation also tested
 - Detected/confirmed issues under high load :
 - File corruption
 - Issue with open-file limit
- * Impressive results for remote data access :
 - Good tool for Hospital queue
 - But impact on network occupancy ?
- * Globally positive but current bugs prevent to go further

Processing from different sites

Derivation Jobs ($\approx 3\text{MB/s}$) - process 500 Events



- * Solution to get rid of local Grid SE without direct access to remote SE
 - Important to run high IO jobs
- * First implementation : BHAM-Birmingham
 - Opportunity : DPM SE decomissioned
 - Local installation :
 - Reason : Bypass request for remote priviledged access on xcache server
 - Monitoring relies on local tool (UK doing development)
 - Production only site running any kind of job accessing Manchester SE (7k slots in MAN vs 400 slot in BHAM)
 - Operational feedback important to ensure minimal manpower support
 - More infos in Mark slides
- * Next implementation : Cambridge (1k slots) accessing QMUL SE
- * If positive : Validated option to transform in diskless sites
 - Operational cost becomes important

- * Available in internal xcache monitoring
 - Traffic
 - Internal activity
- * Under development : Exposing traffic through Grafana pages of ATLAS (DDM dashboard)
 - Panda pilot contains information if file access through xcache
 - Missing : Was file already cached file ? (Never discussed yet with devs)

- * Documented in ATLAS Twiki since not intuitive
- * Integration of xcache server through AGIS update simple and documented
 - Still difference if site hosts Grid SE or not
 - Manuel operation in Rucio necessary

- * Mandatory to solve pending technical issues
- * No country beyond US/UK/DE have publicly committed to contribute to tests over summer
- * Gain from Virtual Placement under evaluation
- * Deployment with Slate :
 - Very appealing : ensure regular updates (security, bugs)
 - Current implementation does not fit with european security rules
 - Initiating discussion on security issues