

t2k.org VO experience

Data Distribution: *What we'd like v.s. -what we've got
Caveat: could be us or the Grid getting it wrong

t2k.org on the GRID

- **T0** KEK
- **T1** RAL (+TRIUMF)
- **T2** IC, IFIC, Lyon, Liverpool, Oxford, PIC, QMUL, Sheffield
- Currently raw data = ~18TB, *only ran in quasi continuous mode*
- Expect 150-350GB/day

Stock Check

- * Simple stock check of t2k.org data - what data is present at each site?
- - Have to trawl LFC to find replicas of files and perform accounting by hand. Misses out dark data.

All site replication report

Generating ALL File Replication Report for Fri Feb 18 06:00:02 2011

lfn:/grid/t2k.org/nd280/raw/ND280/ND280//00003000_00003999

gfe02.grid.hep.ph.ic.ac.uk	: 1234/1234 replicas
srm-t2k.gridpp.rl.ac.uk	: 1234/1234 replicas
t2ksrm.nd280.org	: 1225/1234 replicas
se03.esc.qmul.ac.uk	: 1229/1234 replicas

lfn:/grid/t2k.org/nd280/raw/ND280/ND280//00004000_00004999

gfe02.grid.hep.ph.ic.ac.uk	: 2368/4788 replicas
srm-t2k.gridpp.rl.ac.uk	: 4788/4788 replicas
t2ksrm.nd280.org	: 4257/4788 replicas
se03.esc.qmul.ac.uk	: 4777/4788 replicas

lfn:/grid/t2k.org/nd280/raw/ND280/ND280//00005000_00005999

gfe02.grid.hep.ph.ic.ac.uk	: 1888/1888 replicas
srm-t2k.gridpp.rl.ac.uk	: 1888/1888 replicas
t2ksrm.nd280.org	: 1303/1888 replicas
se03.esc.qmul.ac.uk	: 1056/1888 replicas

lfn:/grid/t2k.org/nd280/raw/ND280/ND280//00006000_00006999

gfe02.grid.hep.ph.ic.ac.uk	: 3078/5451 replicas
srm-t2k.gridpp.rl.ac.uk	: 5444/5451 replicas
t2ksrm.nd280.org	: 7/5451 replicas

lfn:/grid/t2k.org/nd280/raw/ND280/ND280//00007000_00007999

gfe02.grid.hep.ph.ic.ac.uk	: 2193/4484 replicas
srm-t2k.gridpp.rl.ac.uk	: 4395/4484 replicas
t2ksrm.nd280.org	: 4159/4484 replicas

It took 11370.1515441 seconds to generate this report.

T2 Data Management

- * Data management at T2 - where is the storage and how much is accessible?
- - Coupled to inability to easily take stock of data is lack of clarity in information presented by sites (e.g. multiple degenerate entries in `lcg-infosites` se; broadcasting of resources by sites who do not actually support our VO). Not yet implemented space tokens.

FTS for distribution

- * Use FTS to distribute experimental and MC data.
- - No FTS out of T0 (yet), making scripts robust enough to get data from KEK to T1 requires undesirable amount of sophistication (etc). FTS doesn't register to LFC.

RAL LFC

- * Use LFC as single point of contact with data for GRID users
- - LFC interaction flaky during periods of high activity \therefore requires significant amount of users' time to manage registration of processed data from jobs.

Automated Distributed Computing

- * WMS sends jobs to the data
- - Jobs are being sent to the software (ubiquitous), but data is being pulled and pushed all over the place. WMS limitations hamper large-scale distribution (have to cap job submission)