Data access pattern of a Tier2 in ATLAS

Günter Duckeck, Sergey Kalinin
Bergische Universität Wuppertal

Contents

- Short description of T2 at Wuppertal
- Access patterns at Wuppertal and LRZ
- Outlook

T2 at Wuppertal

- Hardware is from HP:
  - WNs: BL220c, 460c, CentOS
  - Storage: DL380G5, 2xRAID6(40TB) pro server, 10 Gb ports
- ~1000 cores(Torque/MAUI), 1 PB storage(dCache)
- New hardware:
  - WN: BL2x220 G7, 386 cores
- T3 at Dortmund use our storage

Sources of information

- Panda(distributed production system) website to get the list of dynamically distributed datasets to Wuppertal.
- dCache logs, no DB interface yet. The logs have detailed transfer decsription(file id, token, etc). A tool has been developed to parse logs and to extract interesting statistics from them. The results for the two last moths are presented in this talk.

The most and the least used datasets

```
ength($1)-5,5); print a"\t\t"$2}'
----- Dataset name                    Accesses ---
ddo.000001...01010                        50019
mc11_7TeV....945_0                        38960
ddo.000001...08010                        36595
data11_7Te...470_0                        32385
mc10_7TeV....044_0                        31386
data11_7Te...510_0                        28849
mc11_7TeV....665_0                        18872
data11_7Te...712_0                        17811
data11_7Te...698_0                        16030
data11_7Te...511_0                        14993
data11_7Te...629_0                        14689
data11_7Te...264_0                        14659
data11_7Te...4_m93                        14405
user.abart...41900                        14296
mc11_7TeV....387_0                        13964
mc10_7TeV....062_0                        13635
20110810__11081                           13503
ddo.000001...05010                        13482
user.mneum...15393                        11677
data11_7Te...189_0                        11568
```

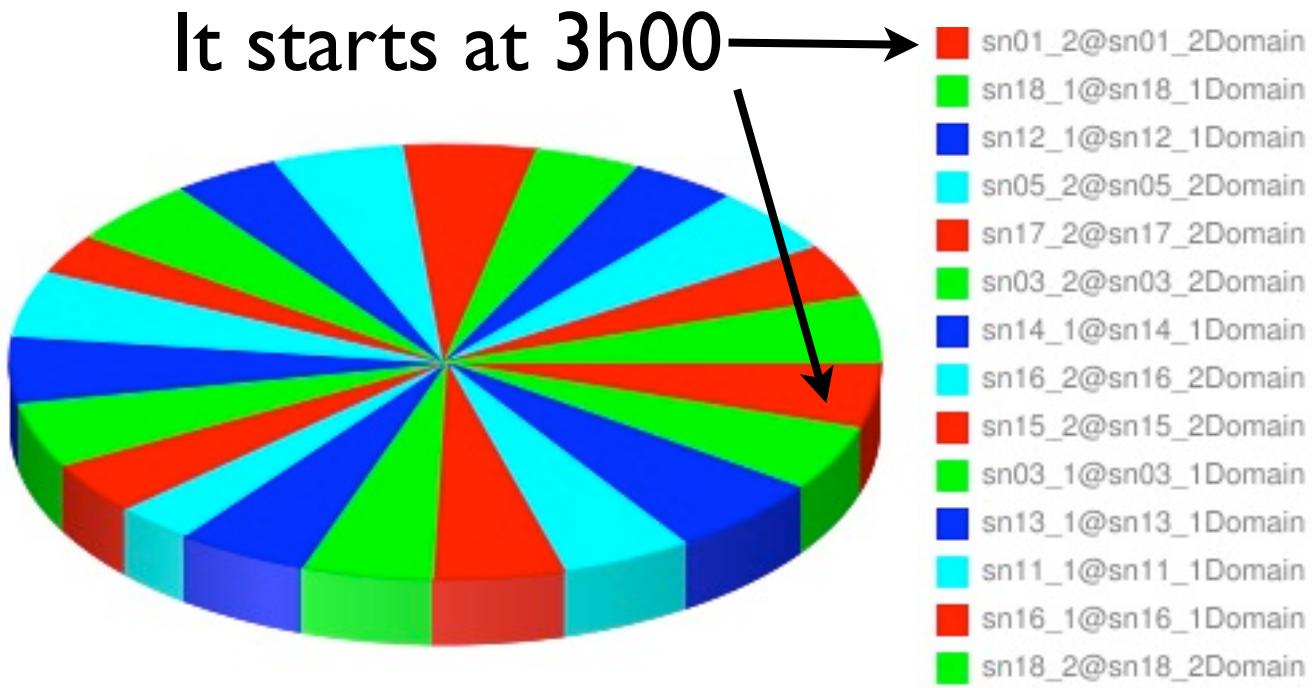
alignment, trigger, cond DB

The tool extracts the information about the number of file accesses. 19 datasets out of 50 the most accessed were dynamically replicated to Wuppertal!

HammerCloud

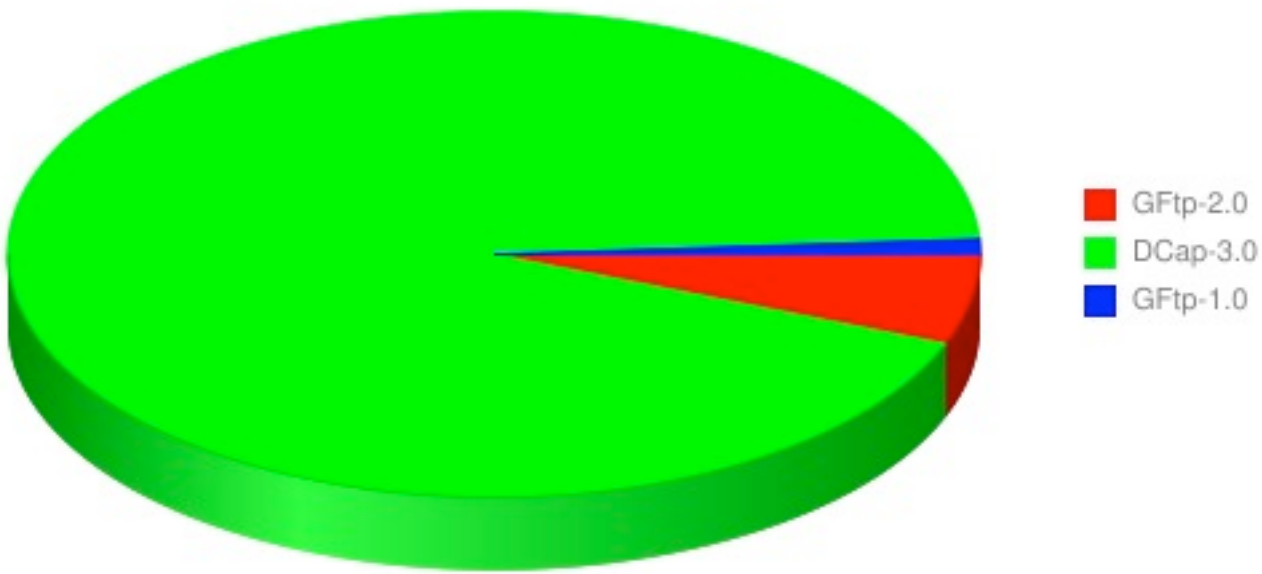
It also returns, of course, the list of datasets which were not used at all and the owner can be asked.

Transfer statistics for pools



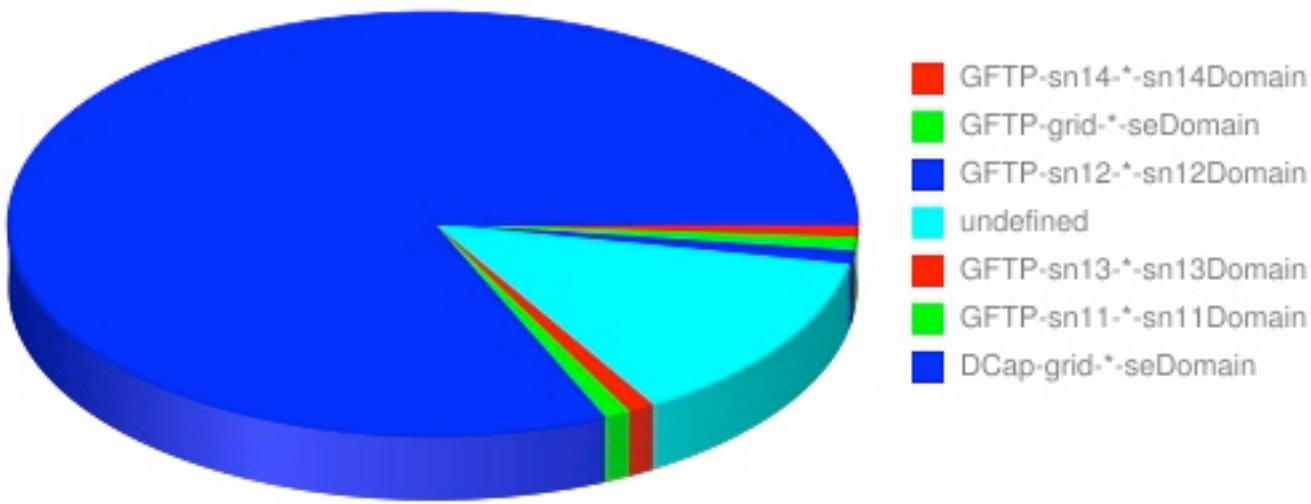
Thanks to our similar hardware,
the distribution is almost perfect.

Transfer statistics for protocols



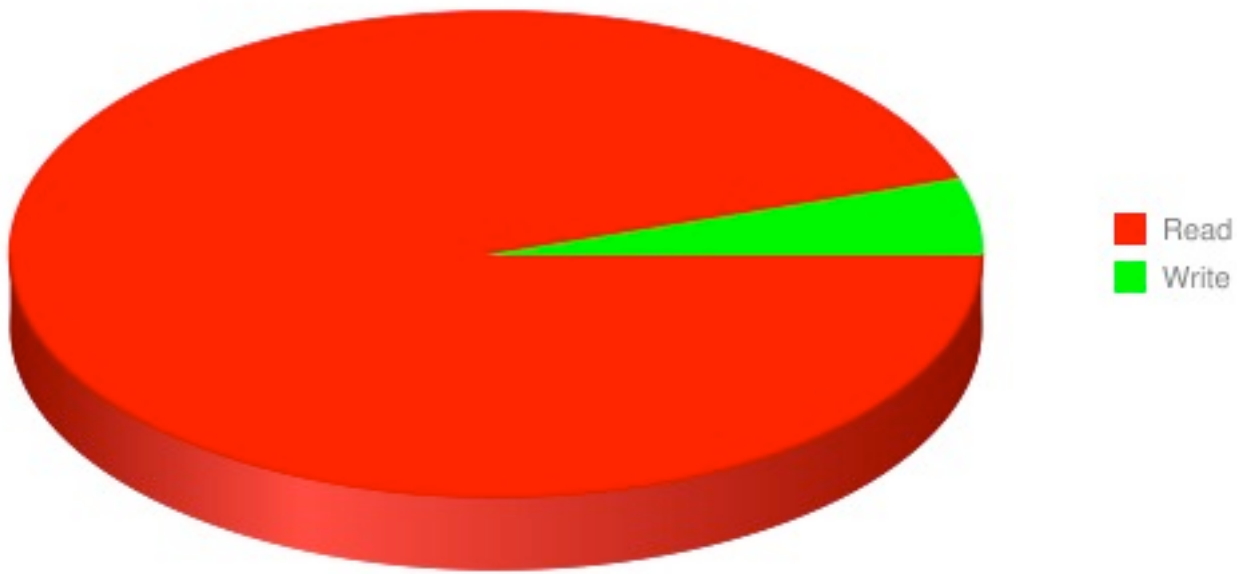
GFtp-1.0 is used by Auger and IceCube. The other ones are by ATLAS.

Transfer statistics for doors



We have 4 GFTP doors and 1 dcap door

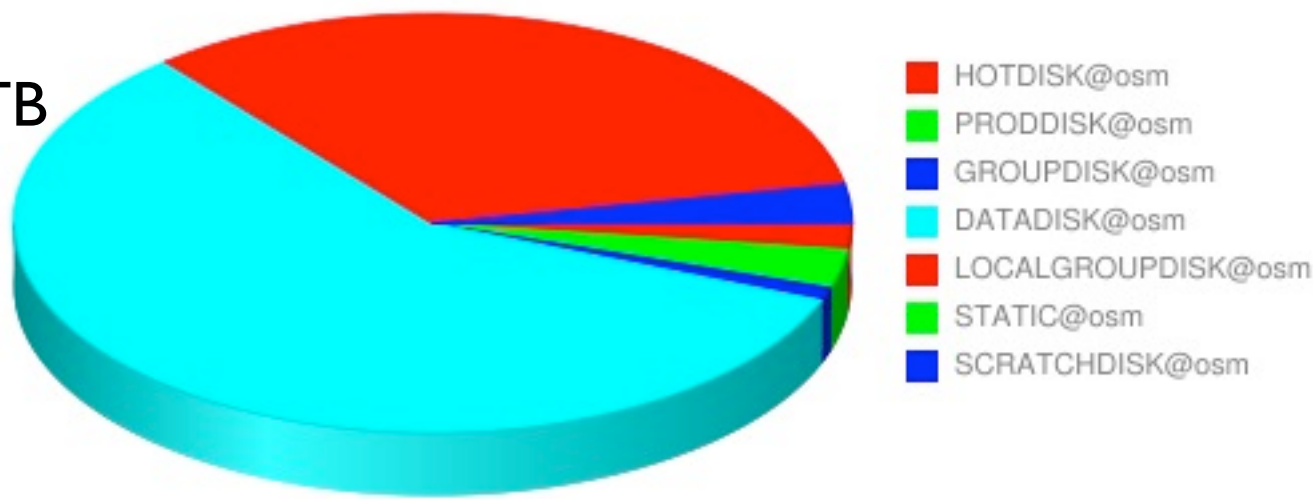
Transfer Read/Writes



95% by volume(79% by number of transfers) are reads!

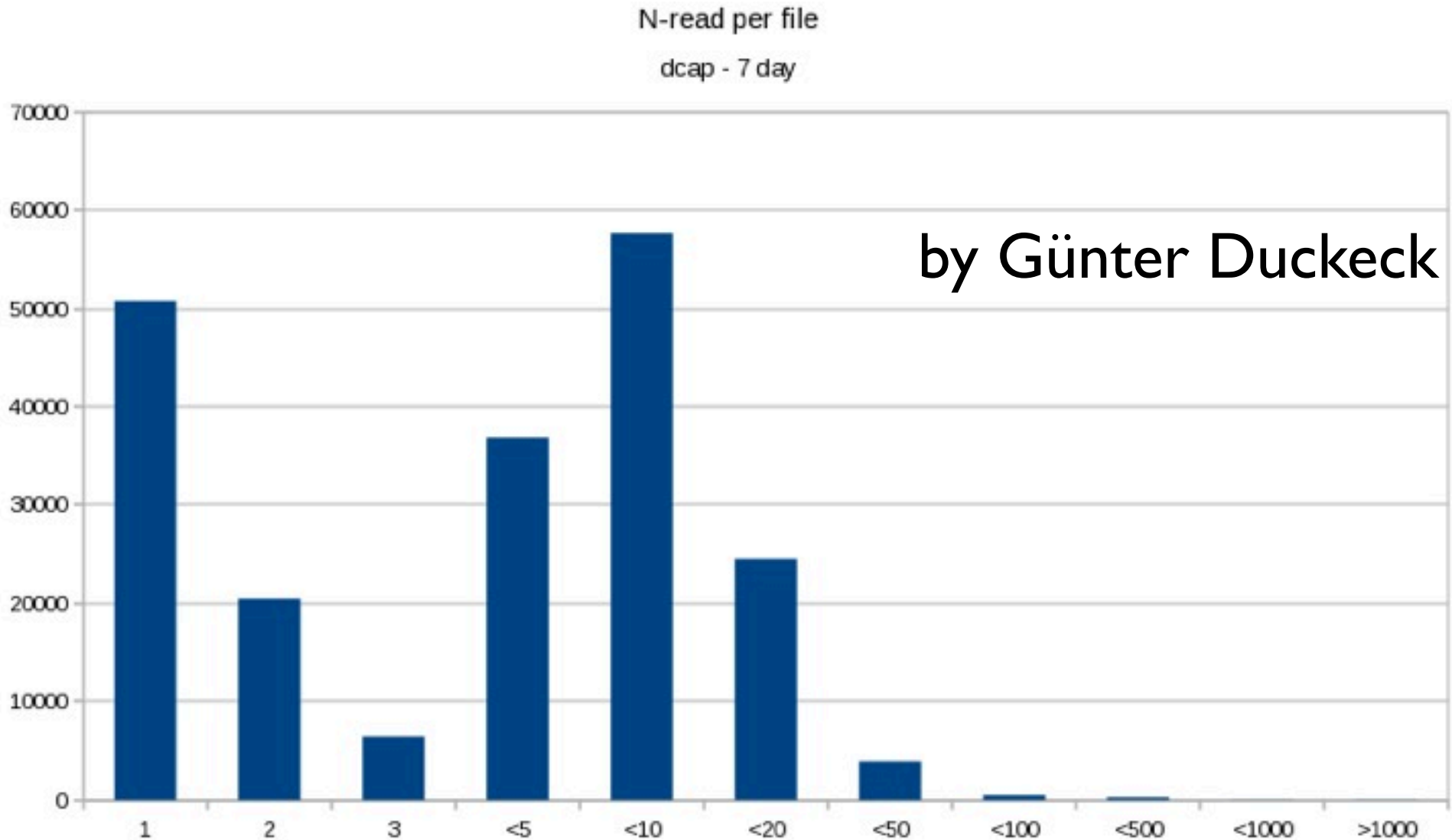
Transfer statistics for space tokens

Used space
DATADISK:471 TB
LOCALGROUPDISK: 118 TB
GROUPDISK: 61 TB



DATADISK is a general purpose token for data and mc.
LOCALGROUPDISK is for local non pledged storage.

File popularity at LRZ



by Günter Duckeck

… nicely pronounced after 1 week …

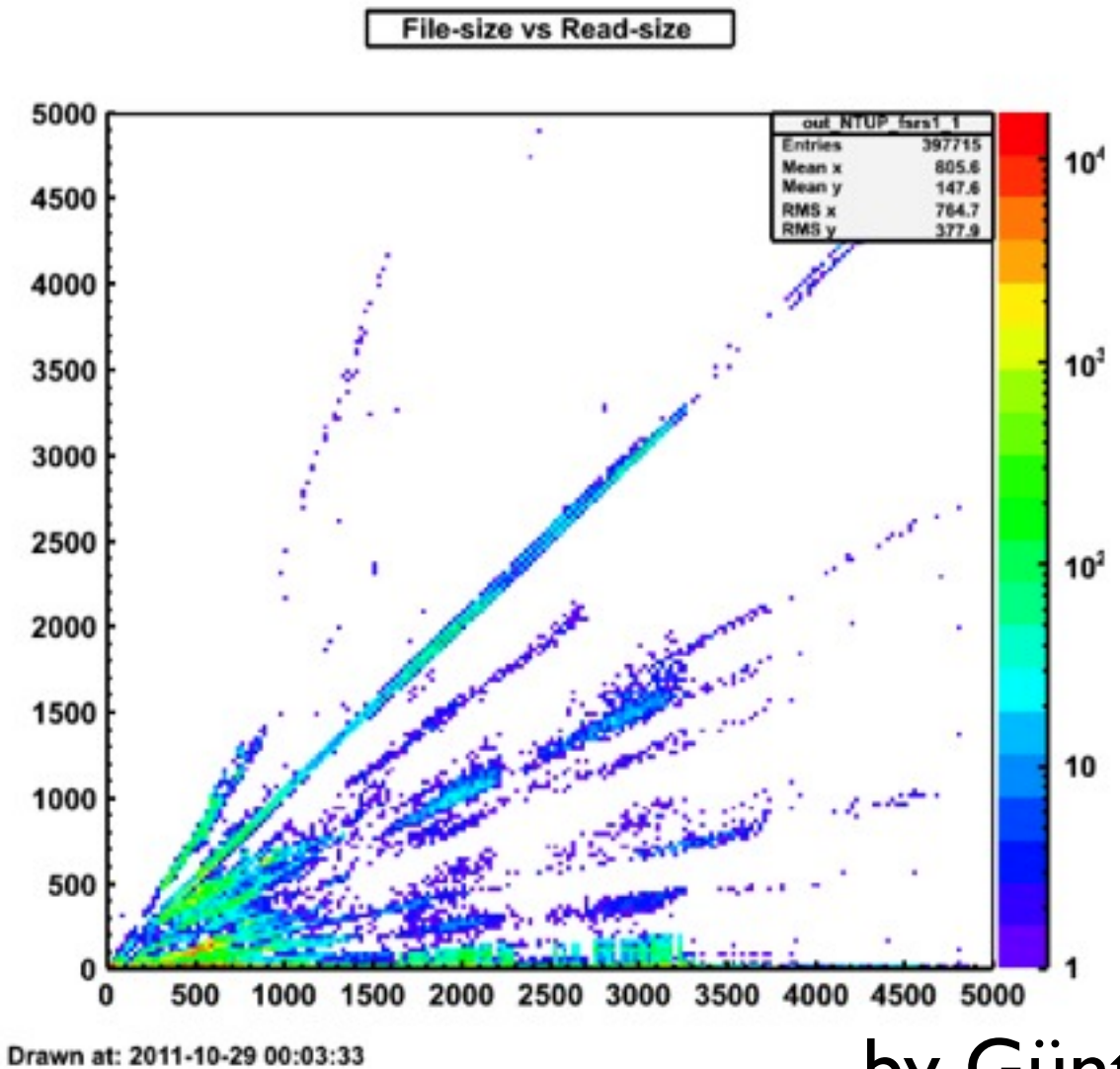
Data volume transferred vs storage accessed

time	Number of different	total Vol (TB)	unique Vol (TB)	unique Vol (files read >
1d	67752	25	15	1
2d	115943	56	24	5
7d	201543	186	52	22
40d	525620	929	120	61

- Only ~20% of data read within 40days (120 TB vs 600 TB)
- For 7 days the ~20 TB most popular files yield ~150 TB of transfer volume
  - Replicating/caching those on fast storage might help a lot for performance

by Günter Duckeck

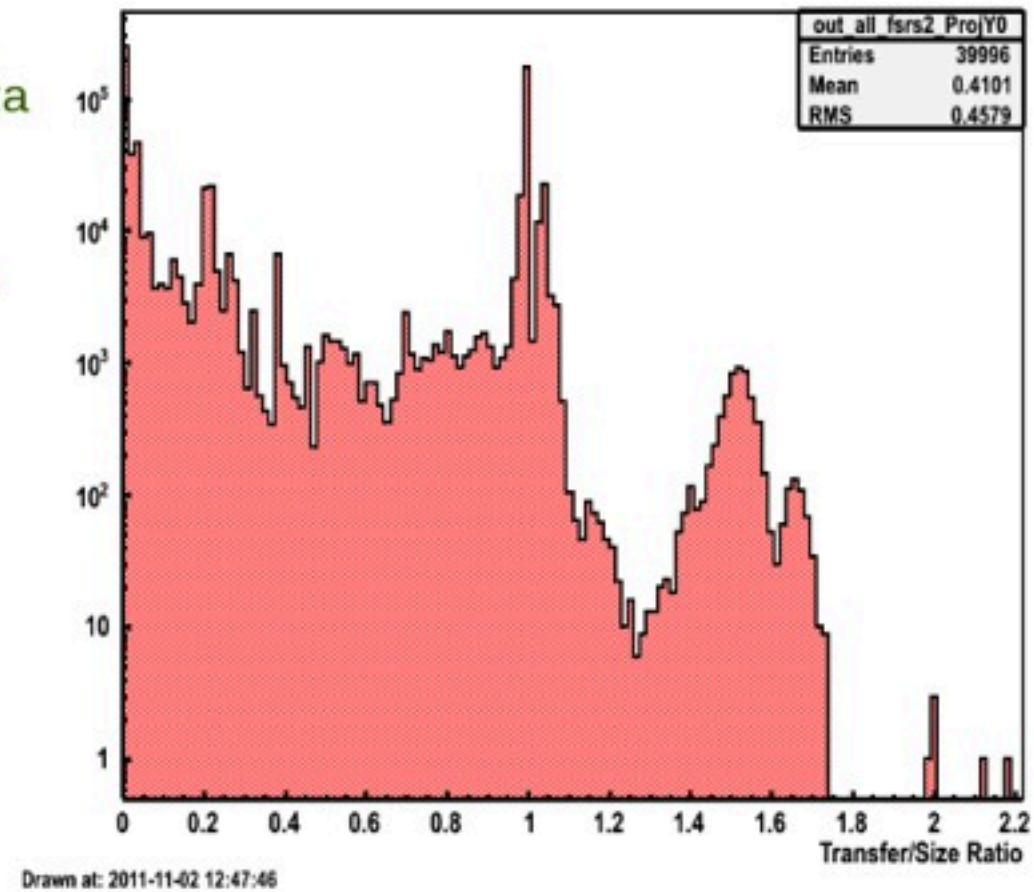
Transfer size vs File size - Ntuple



by Günter Duckeck

Transfer size/File size ratio

- Often for AOD/ntuple not all data is needed
  - Proven by billing stat when comparing transfer size vs file-size
  - Sometimes backward jumps cause repeated reading of blocks → overhead



by Günter Duckeck

Outlook

- A tool has been developed to show real usage of dCache based mass storage for all VOs and tested at Wuppertal
- ATLAS data replication is effective
- There is a room for further developments(AOS/ESD, transfer destination, etc). Can also be combined with other tools(e.g. developed at Munich)