# ATLAS: Experience operating Tier-2 centers

Günter Duckeck, Torsten Harenberg

*Helmholtz Alliance Meeting, December 8th, 2011*

# Content

- DE cloud in ATLAS
- German cloud „particularities“
- Experience running „small“ sites (Tier-3)
- Preparing for real data taking
- Lessons learned from running period with real data
- Summary & Outlook

# ATLAS-wide organization

- ATLAS groups Tier-1/2 (and 3) centers to „clouds“ with one Tier-1 center and associated Tier-2/3s: CA, (CERN), **DE**, ES, FR, IT, ND, NL, TW, UK, US (+ OSG)

- DE cloud:
  - Tier-1: Gridka @ FZK
  - Tier-2: within Germany: DESY-HH, DESY-ZN, Göttingen, LRZ-LMU Munich, MPP Munich, FReiburg and Wuppertal
    outside Germany: CSCS, CYFRONET, Prague, Innsbruck
  - Tier-3 (not all services):  Bonn, Dortmund, Dresden, Mainz, Siegen.

# cloud wide organisation

- weekly operation meetings (operations team: Production (Panda), DDM, Tier-2, Software Installation)
- monthly cloud meeting (all Tier-2s)
- very active mailing list
- full-time Tier-1 contact @ GridKa, participates in GridKa operations meetings, very helpful
- two pilot factories: serving cloud with pilot jobs
- Technical Advisory Board @ GridKa contacts: Günter Duckeck, Torsten Harenberg
- dCache support group

| Row Labels | Sum of jobs | jobs % of total | sum of cpu consumption | cpu consumption % of total | sum of walltime | walltime % of total |
|---|---|---|---|---|---|---|
| CA | 5564091 | 4,51% | 54890849571 | 5,49% | 74796109268 | 5,30% |
| CERN | 5725012 | 4,64% | 25655333608 | 2,57% | 36914974019 | 2,62% |
| DE | 18388207 | 14,90% | 1,59291E+11 | 15,95% | 2,2453E+11 | 15,92% |
| ES | 5592023 | 4,53% | 48528606542 | 4,86% | 72589511741 | 5,15% |
| FR | 14669447 | 11,89% | 1,16257E+11 | 11,64% | 1,78451E+11 | 12,65% |
| IT | 5279660 | 4,28% | 48231371137 | 4,83% | 67949502617 | 4,82% |
| ND | 5736541 | 4,65% | 67479155643 | 6,75% | 78878185065 | 5,59% |
| NL | 8039694 | 6,51% | 68538494413 | 6,86% | 96968445523 | 6,87% |
| TW | 2820073 | 2,29% | 19807322060 | 1,98% | 28375676490 | 2,01% |
| UK | 11803730 | 9,56% | 1,39155E+11 | 13,93% | 2,04041E+11 | 14,47% |
| US | 39787012 | 32,24% | 2,51144E+11 | 25,14% | 3,46982E+11 | 24,60% |
| **Grand Total** | **123405490** | **100,00%** | **9,98979E+11** | **100,00%** | **1,41048E+12** | **100,00%** |

01-01-2010 - 31-12-2010: DE cloud offers ~16% of CPU time to ATLAS

**→ 2nd largest cloud in ATLAS**

# „German particularities“

- all (full, up to Tier-2) sites in Germany use dCache exclusively.
  - ideal for dCache.org (testing!), ideal for sites (support!)
  - cloud developed accounting and monitoring tools (see also Sergey‘s talk)
  - Gridlab @ DESY-HH
  - personal is very active in german support group + workshops
  - Happyface plugins
  - planned: NFSv4 (aka pNFS) , redirector (see Outlook)

# integrating „small“ (Tier-3) sites

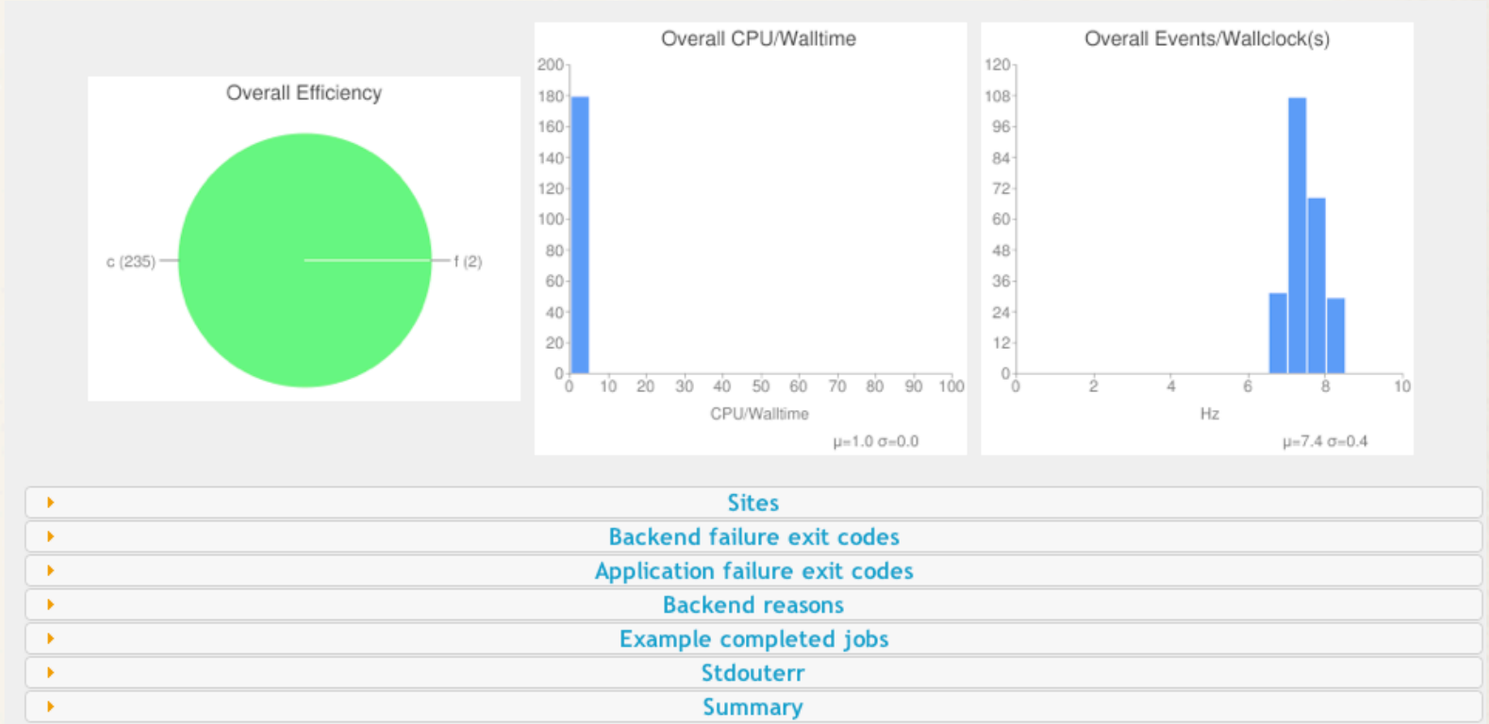
- Realized three modes of operation for small sites:
  - Bonn/Siegen/Mainz: *Data-only*: sites install Storage Element and can use all ATLAS data handling tools (distributed data management - DDM)
  - Dresden: *„full“* site
  - Dortmund: (first time in ATLAS) *CPU-only* site, uses Wuppertal Storage (special setup in Wuppertal dCache required), maybe prototype for cloud (a la EC2) usage

# preparing for real data

- although sites worked well in „pre-data“ times, „real“ test with a typical amount of user analysis jobs needed → „HammerCloud“

- Can run highly configurable, „real“ analysis jobs

- Extremely helpful to find bottlenecks and „not-so-optimal“ settings




meanwhile also/mainly used 24/7 as „functional“ test to continuously check that sites can perform user analysis

# first running period with real data

- data challange: adaptions performed
  - ATLAS data model was strictly hierarchical: all data went through Gridka. Problems with disk space at GridKa.
  - changed recently: „Big Tier-2“ (W, FR, DESYs, LMU, MPP, GÖ) can get data directly from other sites using several new FTS channels, reduces (temp.) disk space requirement at GridKa.
    Network: HEPPI??
  - Firewall at GridKa: saturated by traffic - known IP ranges of associated Tier-2s excluded from firewall

# Data transfers

Data export:

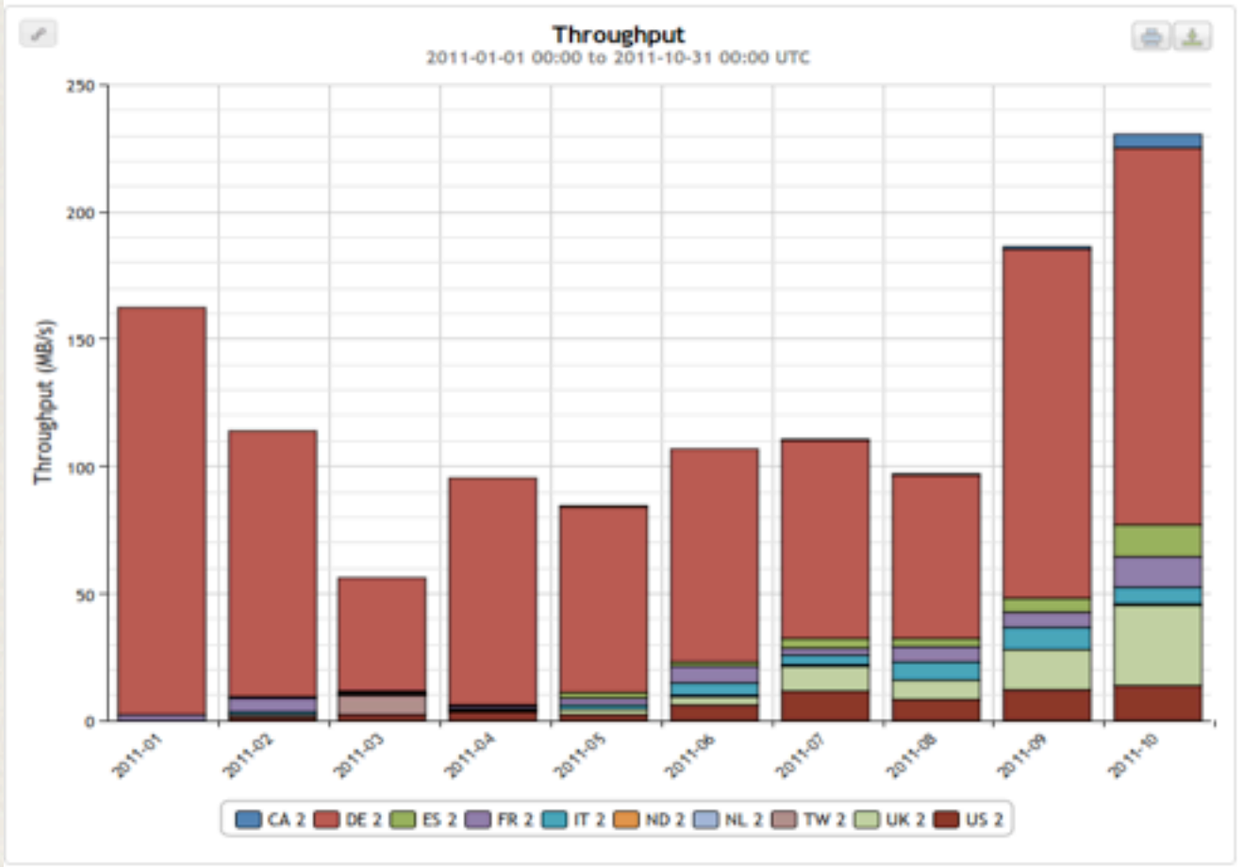
GridKa → Tier2 Sites

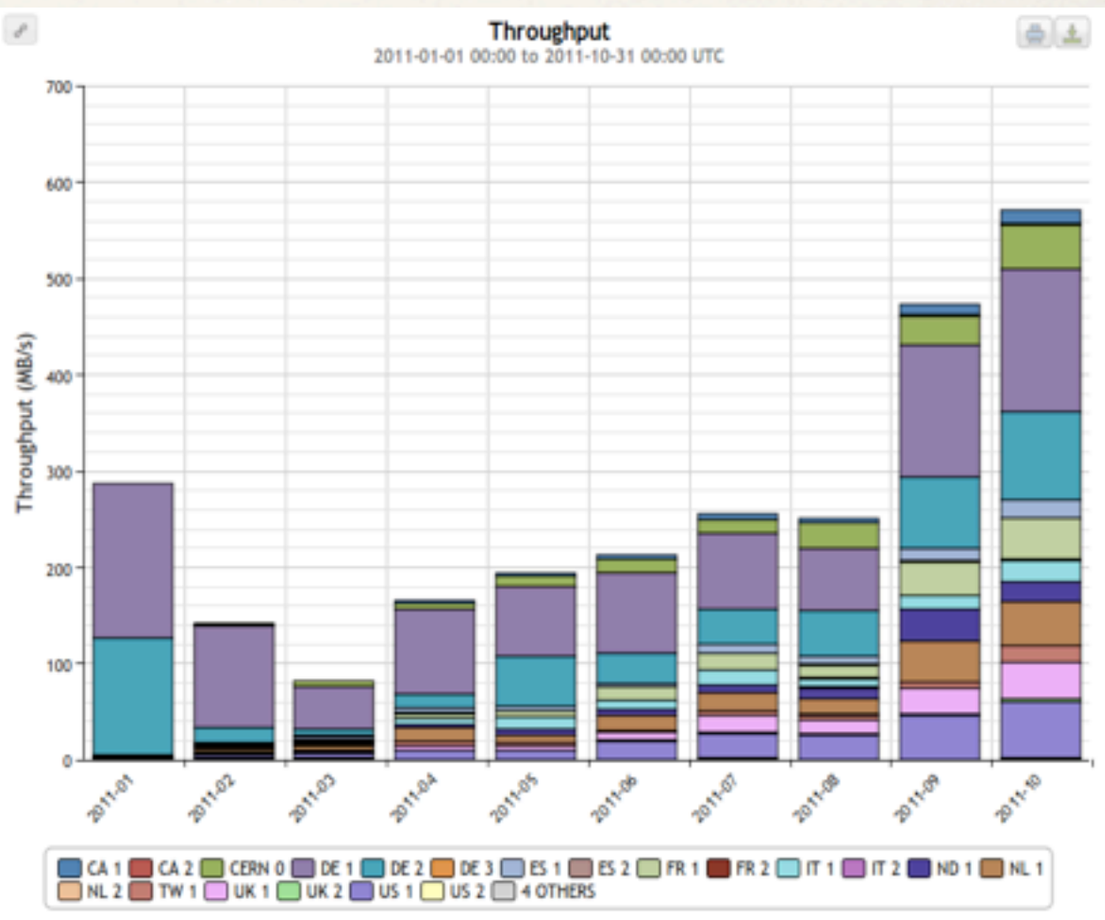
- direct transfer Non-DE T2 since June
- DE T2 dominant

data import:

to DE Tier2 Sites

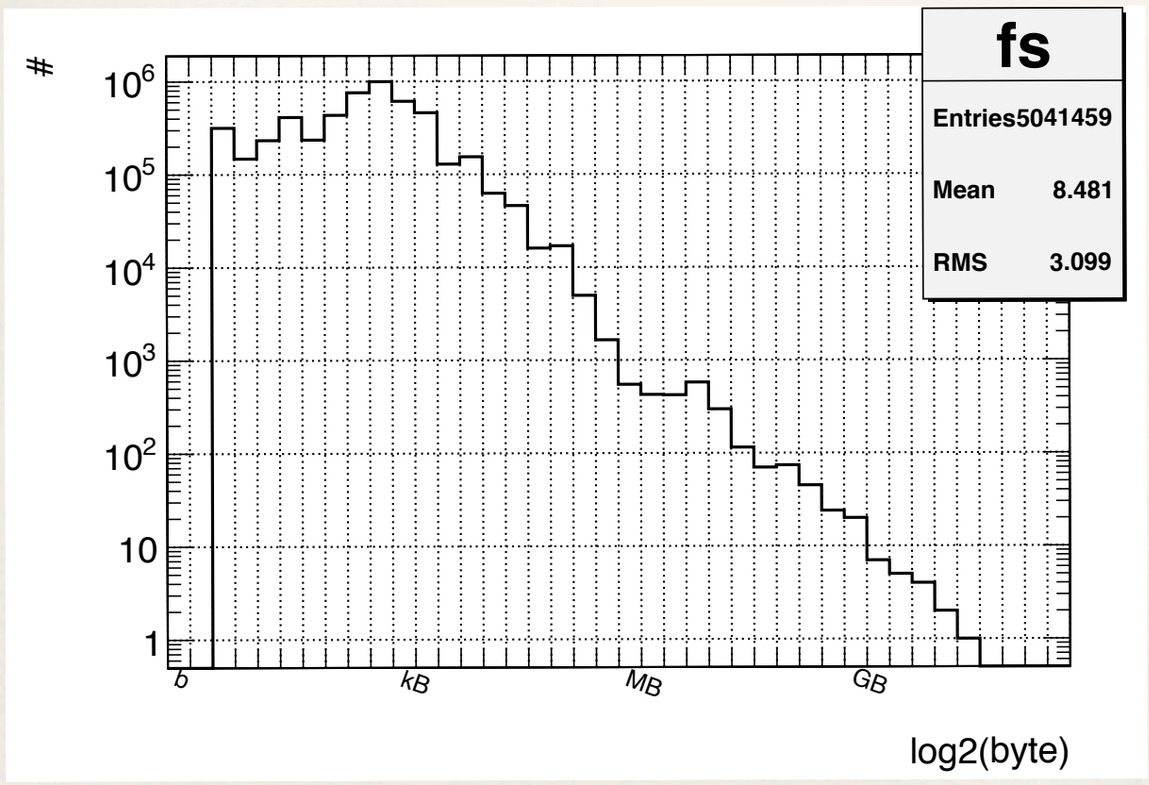
- direct transfers from non-DE T2 since June
- transfers from outside cloud dominate

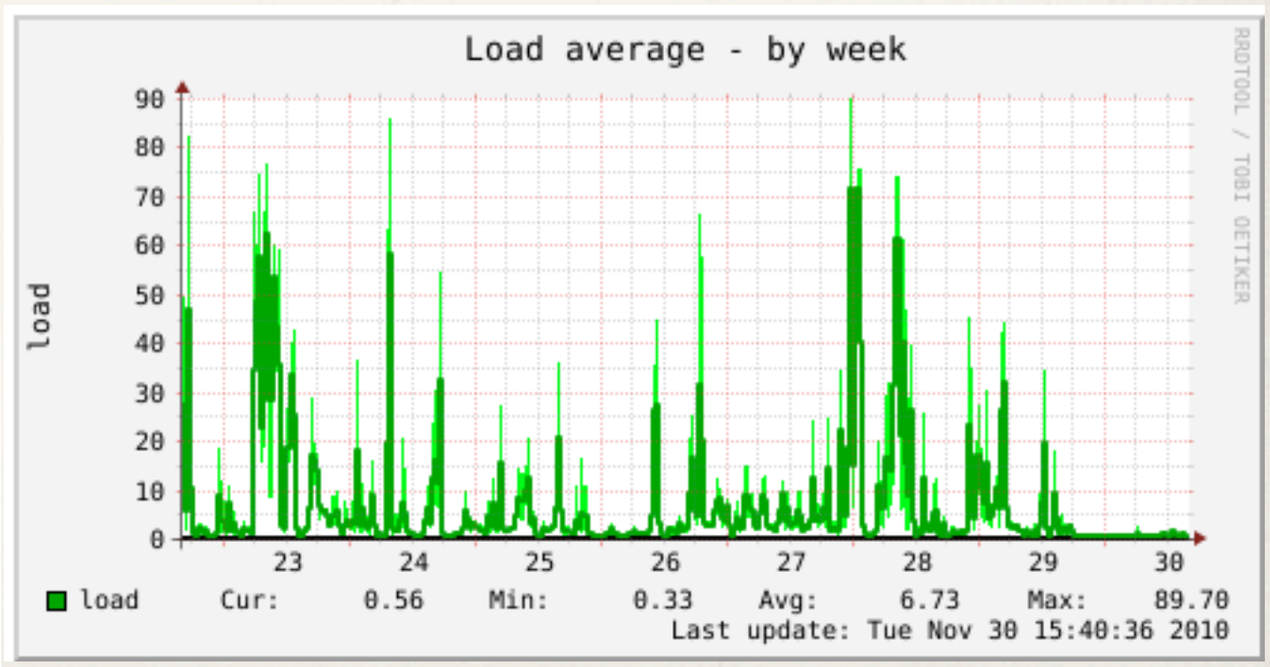


# more jobs - software distribution

- Classical approach to distribute ATLAS software: special jobs write software into (NFS-)shared space.
- ATLAS software consists of 100.000s of small files - NFS or Lustre under high load → solution: CVMFS
- meanwhile ATLAS-recommendation, also esp. useful for Tier-3s



# the „death“ of monitoring

- monitoring exists on several (too many?) levels (see also talk from O. Oberst)
- often a quick overview needed („is the cloud healthy?“)
- → Happyfaces

**Cloud Monitoring 2.0** Tuesday, 06 December, 14:00:02 (CET) Home Storage Usage Cloud software Help

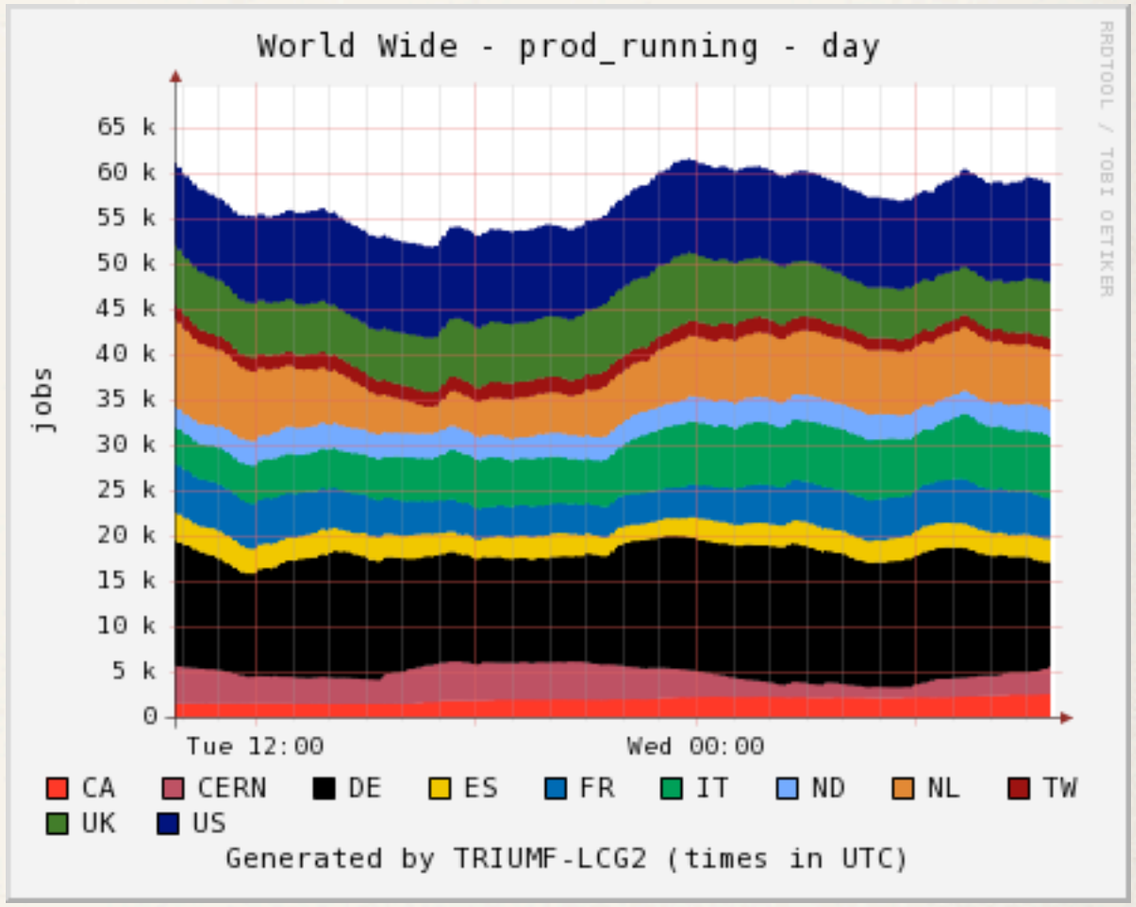
| *Cloud: DE* | Panda Analysis | | | Panda Production | | | ATLAS SAM | | OPS SAM Daily Month Year | | | Availability | | | Site Exclusion | | GGUS | | Downtimes | | Ganga Robot | | SW |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Site | Act | Run | Eff% | Act | Run | Eff% | CE | SRMv2 | CE | sBDII | SRMv2 | Day | Dec | Nov | Panda Status | DDM Status | Open | Closed | Now | Soon | PANDA | LCG | |
| CSCS-LCG2 | 696 | 354 | 37 | 938 | 101 | 77 | ok | ok | ok | ok | ok | unk | 100 | 91 | online | online | 0 | 77013 | UP | No | no-test | 0 | ok |
| CYFRONET-LCG2 | 1304 | 994 | 84 | 2602 | 590 | 100 | ok | ok | ok | ok | ok | unk | 100 | 97 | online | online | 0 | 2 | UP | No | no-test | ok | ok |
| DESY-HH | 657 | 636 | 57 | 747 | 1017 | 98 | ok | ok | ok | ok | ok | unk | 94 | 100 | online | online | 76885 | 0 | UP | No | no-test | ok | ok |
| DESY-ZN | 234 | 332 | 25 | 504 | 502 | 85 | ok | ok | ok | ok | ok | unk | 100 | 100 | online | online | 0 | 76068 | UP | No | no-test | ok | ok |
| FZK-LCG2 | 1610 | 634 | 69 | 9895 | 4409 | 100 | ok | ok | ok | ok | ok | unk | 100 | 99 | online | blacklisted | 2 | 3 | UP | No | no-test | n/a | ok |
| GoeGrid | 9 | 369 | 57 | 6 | 25 | 0 | ok | ok | ok | ok | ok | unk | 83 | 81 | online | online | 77063 | 4 | UP | No | no-test | no-test | ok |
| HEPHY-UIBK | 10 | 1 | 0 | 109 | 0 | 0 | ok | ok | F | ok | ok | unk | 83 | 73 | online | online | 0 | 76086 | UP | No | no-test | blacklisted | ok |
| LRZ-LMU | 0 | 896 | 94 | 0 | 391 | 99 | ok | ok | ok | ok | ok | unk | 100 | 85 | online | online | 0 | 0 | UP | No | no-test | ok | ok |
| MPPMU | 1416 | 81 | 86 | 1002 | 1228 | 99 | ok | ok | ok | ok | ok | unk | 69 | 93 | online | online | 0 | 4 | UP | No | no-test | ok | ok |
| praguelcg2 | 0 | 0 | 0 | 60 | 5 | 100 | ok | ok | ok | ok | ok | unk | 100 | 88 | online | online | 0 | 3 | UP | No | no-test | no-test | ok |
| PSNC | 0 | 0 | 0 | 0 | 109 | 73 | ok | ok | ok | ok | ok | unk | 100 | 87 | online | online | 0 | 2 | UP | No | no-test | ok | ok |
| UNI-FREIBURG | 718 | 692 | 53 | 1361 | 268 | 96 | ok | ok | ok | ok | ok | unk | 100 | 93 | online | online | 0 | 2 | UP | No | no-test | ok | ok |
| wuppertalprod | 49 | 388 | 92 | 1589 | 466 | 100 | ok | ok | ok | ok | ok | unk | 100 | 99 | online | online | 0 | 0 | UP | No | no-test | no-test | ok |

**Storage Usage & Data Transfers**

| | dCache/DPM | SPACE TOKENS (TB) | | | | | Data Transfer | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Site | Version | DATADISK | GROUPDISK | PRODDISK | SCRATCHDISK | LOCALGROUPDISK | DATADISK | GROUPDISK | PRODDISK | SCRATCHDISK | LOCALGROUPDISK |
| CSCS-LCG2 | 1.9.5-27 | 349/422 | 16/49 | 1/10 | 20/27 | 0/10 | 98.1 | 0.0 | 100.0 | 0.0 | 0.0 |
| CYFRONET-LCG2 | 1.8.2 | 182/236 | 29/54 | 6/11 | 14/27 | 7/11 | 100.0 | 0.0 | 100.0 | 0.0 | 0.0 |
| DESY-HH | 1.9.12-12 | 470/551 | 313/363 | 11/27 | 45/68 | 260/428 | 99.3 | 0.0 | 100.0 | 100.0 | 97.6 |
| DESY-ZN | 1.9.12-12 | 245/293 | 44/117 | 6/10 | 26/35 | 229/293 | 99.8 | 0.0 | 99.7 | 0.0 | 0.0 |

# Summary and Outlook

- DE cloud perfoms pretty well
- 2nd largest (and one of the most stable) cloud, usage near 100%
- good communication, very active and efficient operations team
- Tier-1 contact more than helpful
- German sites are „dCache only“



# Outlook: dCache activities

- more monitoring and accounting (see Sergey‘s talk)
- developing Redirector and integrating/attaching it to dCache and the ATLAS DDM
- continue good support for German sites (+ workshops)
- testing / prototyping to NFSv4.1 / pNFS

# Outlook: data analysis optimizations

- Grid analysis jobs
  - improve failure rate and turnaround = wait time for last job
- direct-IO/ROOT-IO optimization - TTreeCache
- more performant & robust IO protocol dcap → xrootd, nfs4.1 ??
- refer to data access analysis
  - balanced data distribution over pools
  - additional caching for most popular files