Computing for HEP in the Czech Republic

Jiří Chudoba

3.4.2007

Institute of Physics, AS CR, Prague

Outline

- Groups involved
- Resources
  - Network
  - Hardware
  - People
- Usage
- Some technical issues
- Outlook

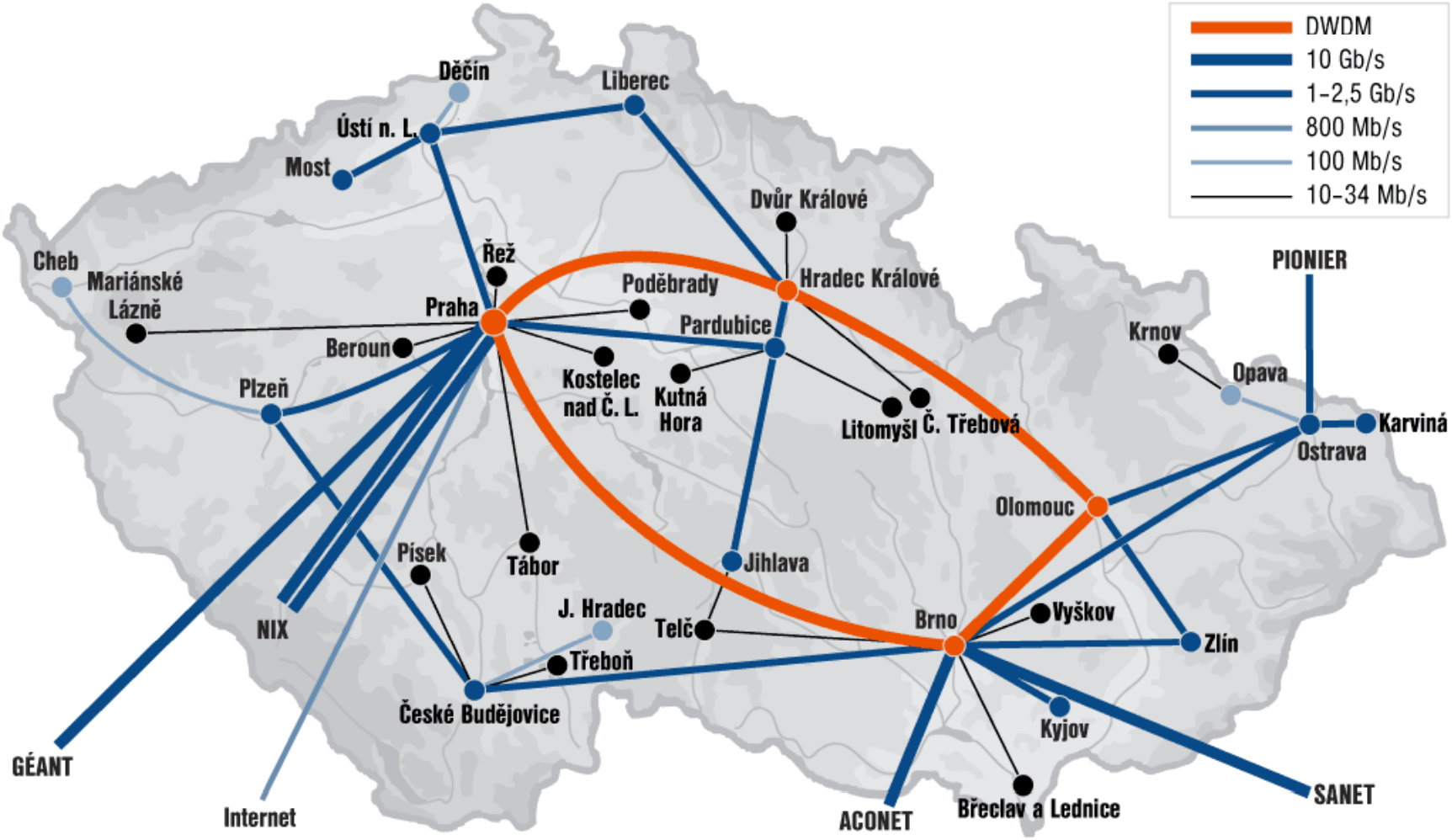
HEP Groups in CZ with Major Computing Requirements

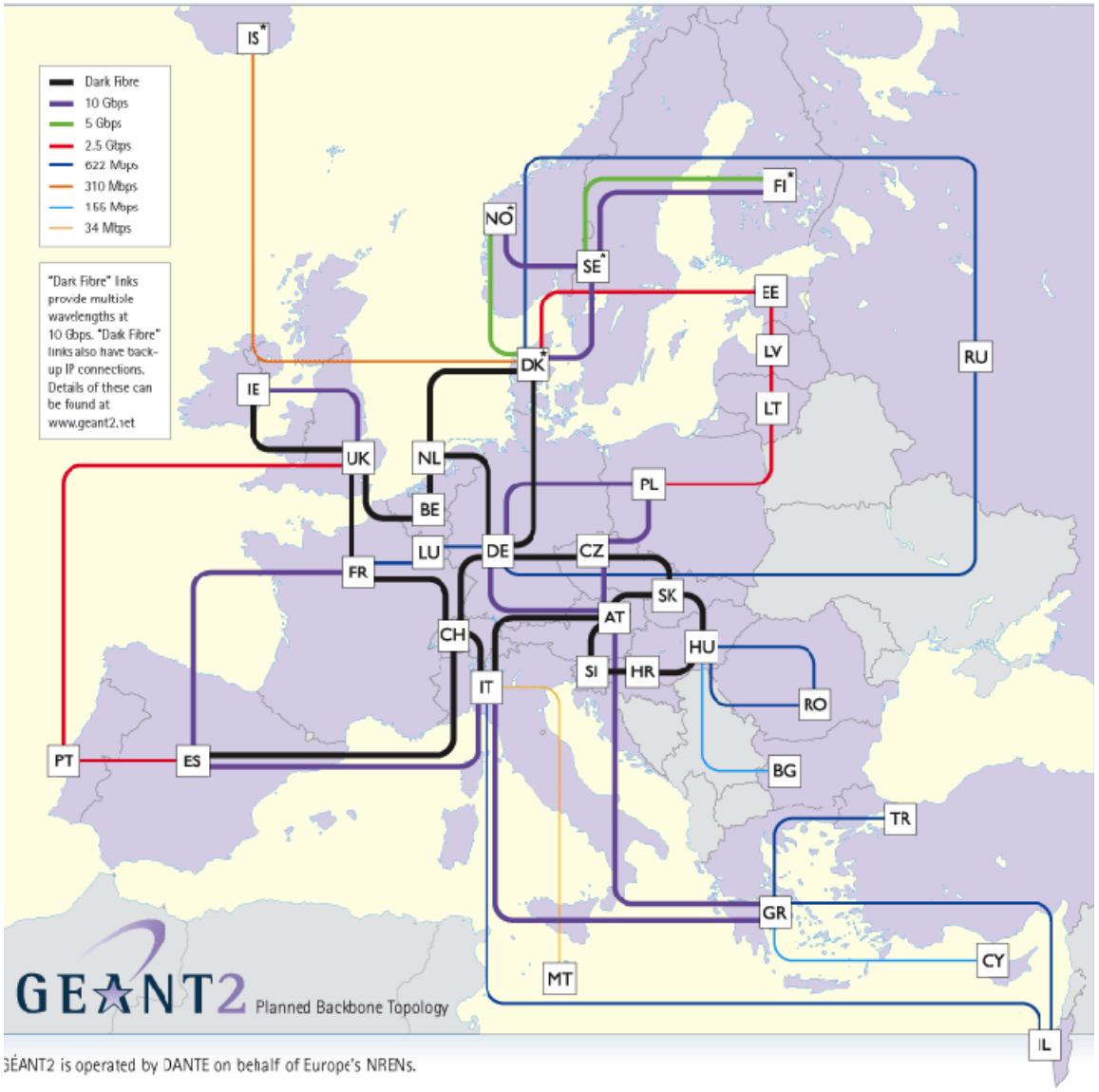
- Institute of Physics, AS CR, Prague
  - ATLAS, ALICE, D0, H1, CALICE
- Institute of Nuclear Physics, AS CR, Řež u Prahy
  - ALICE, STAR
- Czech Technical University, Prague
  - ALICE, ATLAS, STAR, COMPASS
- Charles University, Faculty of Mathematics and Physics, Prague
  - ATLAS, NA49, COMPASS

Network Connection

- provided by CESNET
  - Czech National Research and Education Network provider
  - Association of legal entities formed by universities and the Academy of Sciences of the CR

CESNET2 Topology

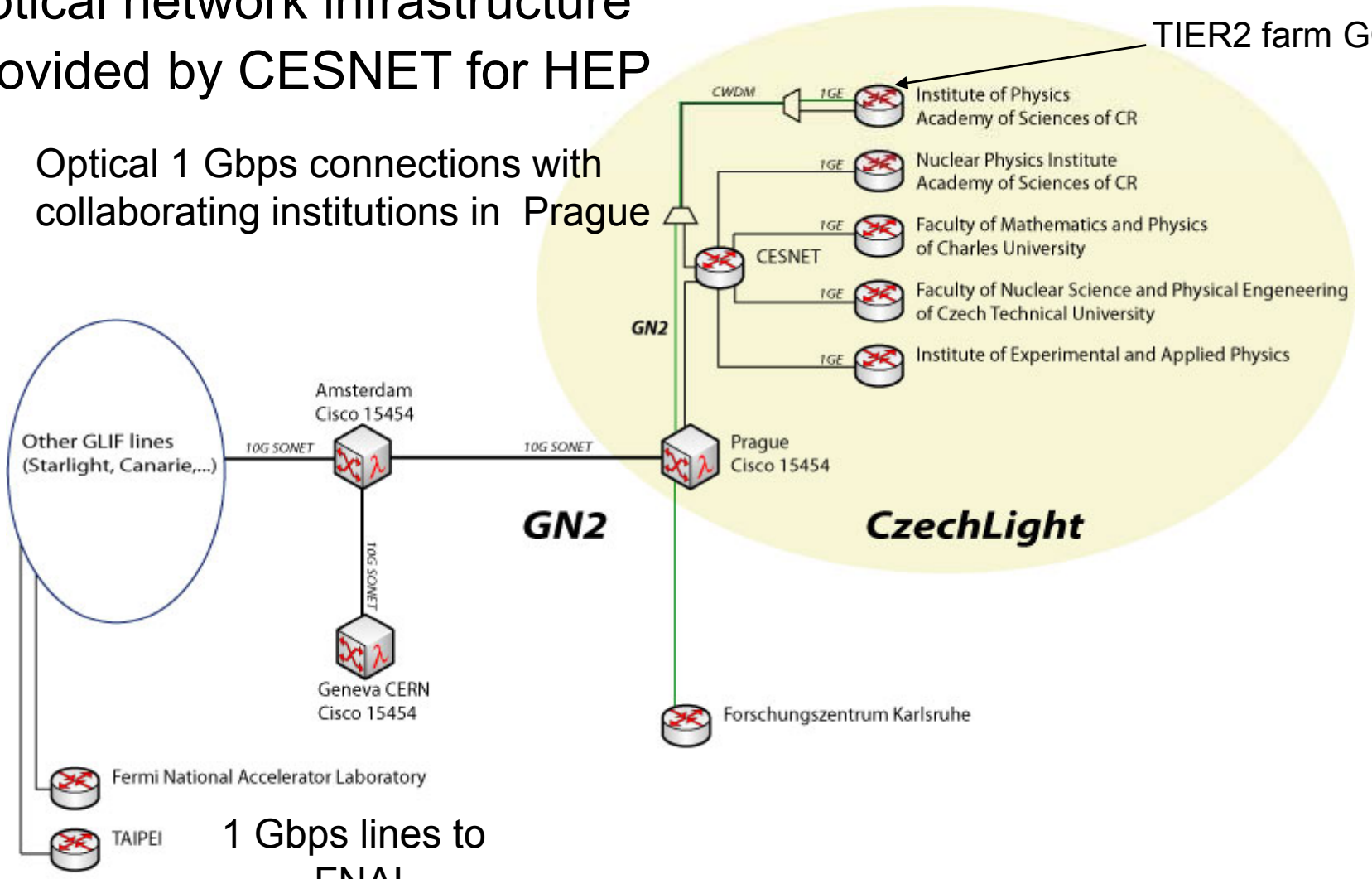




GEANT2 planned topology

Optical network infrastructure Provided by CESNET for HEP

Optical 1 Gbps connections with collaborating institutions in Prague



1 Gbps lines to
- -FNAL
- -TAIPEI
- -FZK

13.9.2006
http://www.particle.cz/farm/czechlight/dokumenty/CL_net.pdf

Hardware Resources

- Tier2 centre (farm GOLIAS) in the Regional Computing Centre for Particle Physics in the Institute of Physics
  - facility shared with non HEP projects
  - contributions from Institute of Nuclear Physics
- CESNET farm SKURUT
  - provided for EGEE project, used by ATLAS, D0, Auger and others
- Smaller local clusters
  - for analysis
  - not shared via grid

Prague T2 center

- GOLIAS at FZU
  - Heterogeneous hardware, due to irregular HW upgrades (from 2002)
  - 140 nodes (including Blade servers), 320 cores
  - ~400k SpecInt2000

GOLIAS

Prague T2 center

- 55 TB of raw disk space;
  4 disk arrays
  - 1 TB array (RAID 5), SCSI disks
  - 10 TB array (RAID 5), ATA disks
  - 30 TB array (RAID 6), SATA disks
  - 14 TB iSCSI array with SATA disks
- PBSPro server, fairshare job ratios

D0	ATLAS	ALICE+STAR	Others
50	19	7	23

- possible usage of solid-state physics group resources
- Network Connections
  1 Gbps lines
  - GEANT2 connection
  - FNAL, p2p, CzechLight/GLIF
  - Taipei, p2p, CzechLight/GLIF
  - FZK Karlsruhe, p2p over GN2, from 1 September 2006
- 150 kW total electric power for computing, UPS, Diesel
- 10 Gbps router in 2007

CESNET farm skurut

- only a fraction for EGEE computing
  - 30 dual nodes, used for ATLAS, AUGER, D0 and non HEP projects (VOCE)
  - OpenPBS
  - also used as a testbed for mw, for virtualisation
  - prototype connection to the METAcentre resources (a grid-like CESNET initiative)

EGEE project

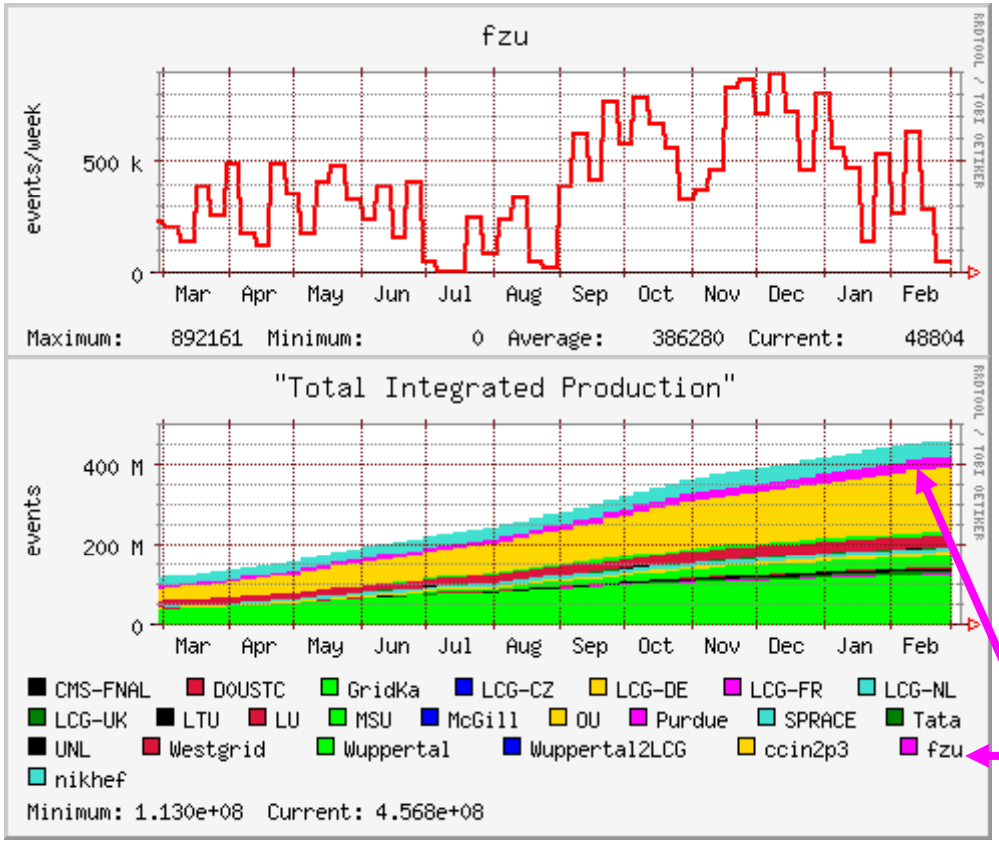
- CESNET is the Czech partner since EDG
- Current involvement
  - JRA1 - development, CZ-IT cluster
    - LB, JP
  - SA1 - operations, monitoring, user support (TPM), preproduction and preview testbeds
  - NA4 - computational chemistry, general consulting, porting of applications

Manpower

- scarce resource
  - 4 administrators for GOLIAS (not 4 FTEs) in charge of HW, OS, MW, network, monitor, support, ...
- local administration often as a part time job for students

D0

- participation in MC production and data reprocessing



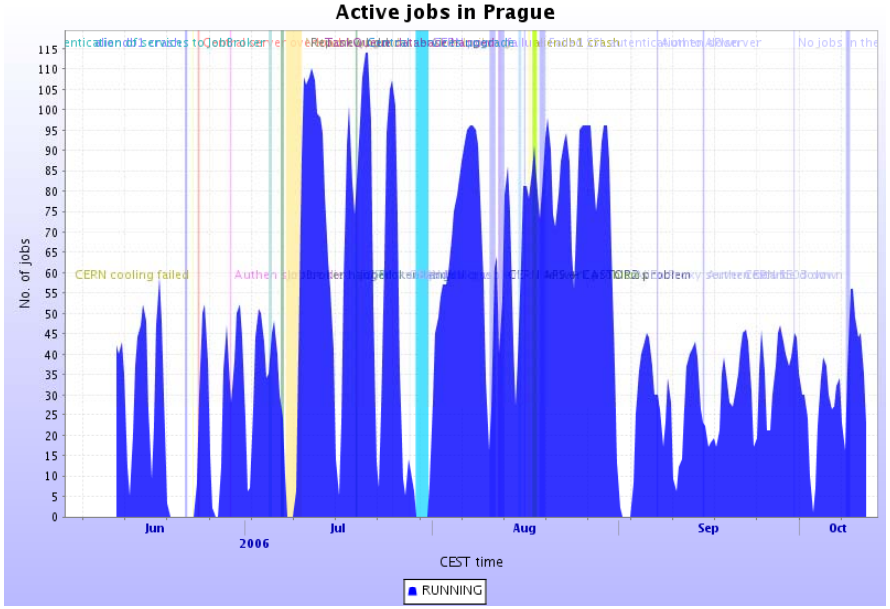
MC production in 2006:

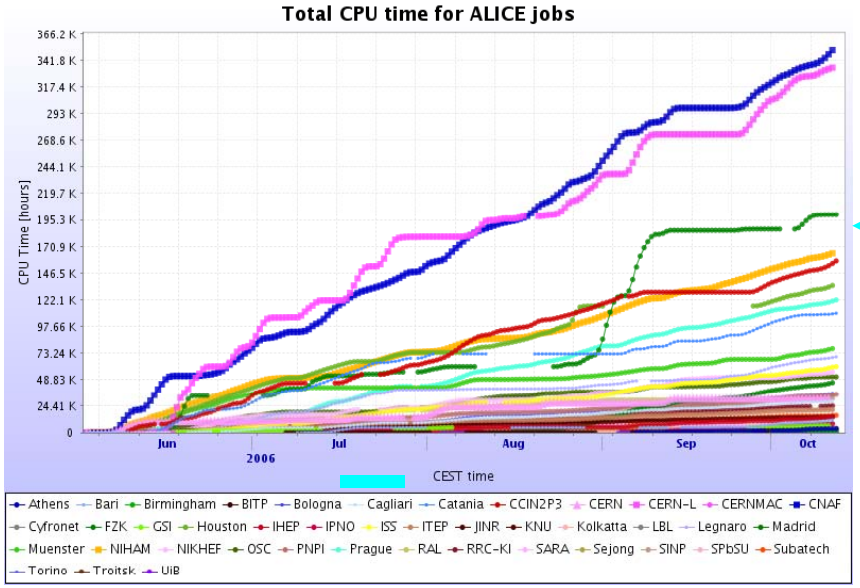
- 17.5 mil. events generated
- 35000 jobs
- 1.27 TB copied to FNAL
- 5 % of total production

Reprocessing
- 26 mil. events
- 2 TB

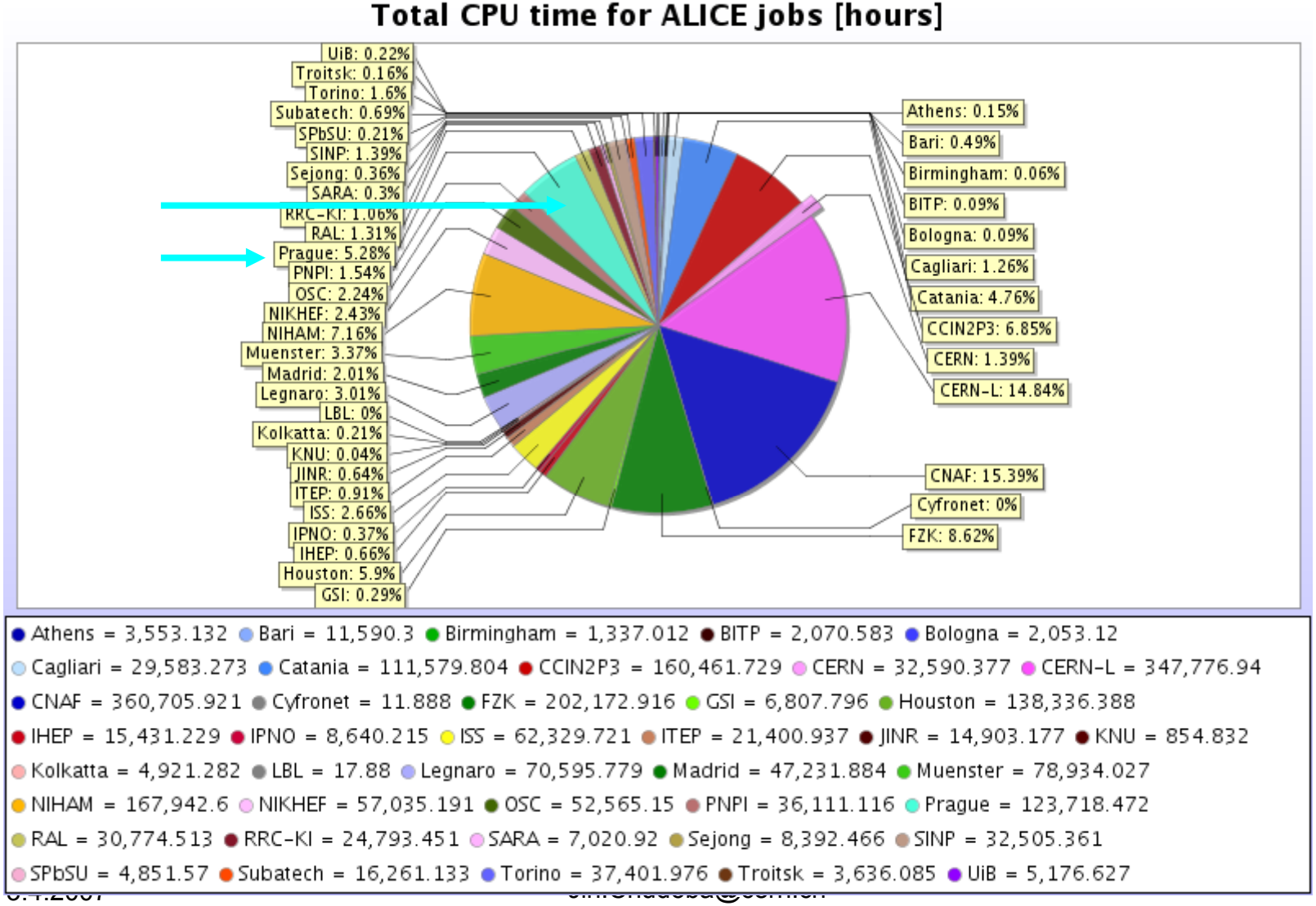
ALICE production in 2006

- running services:
  - vobox
  - LFC catalogue (not used)
  - SE in Rez
- production started in June 2006
- ALICE was able to use idle resources otherwise reserved for other experiments

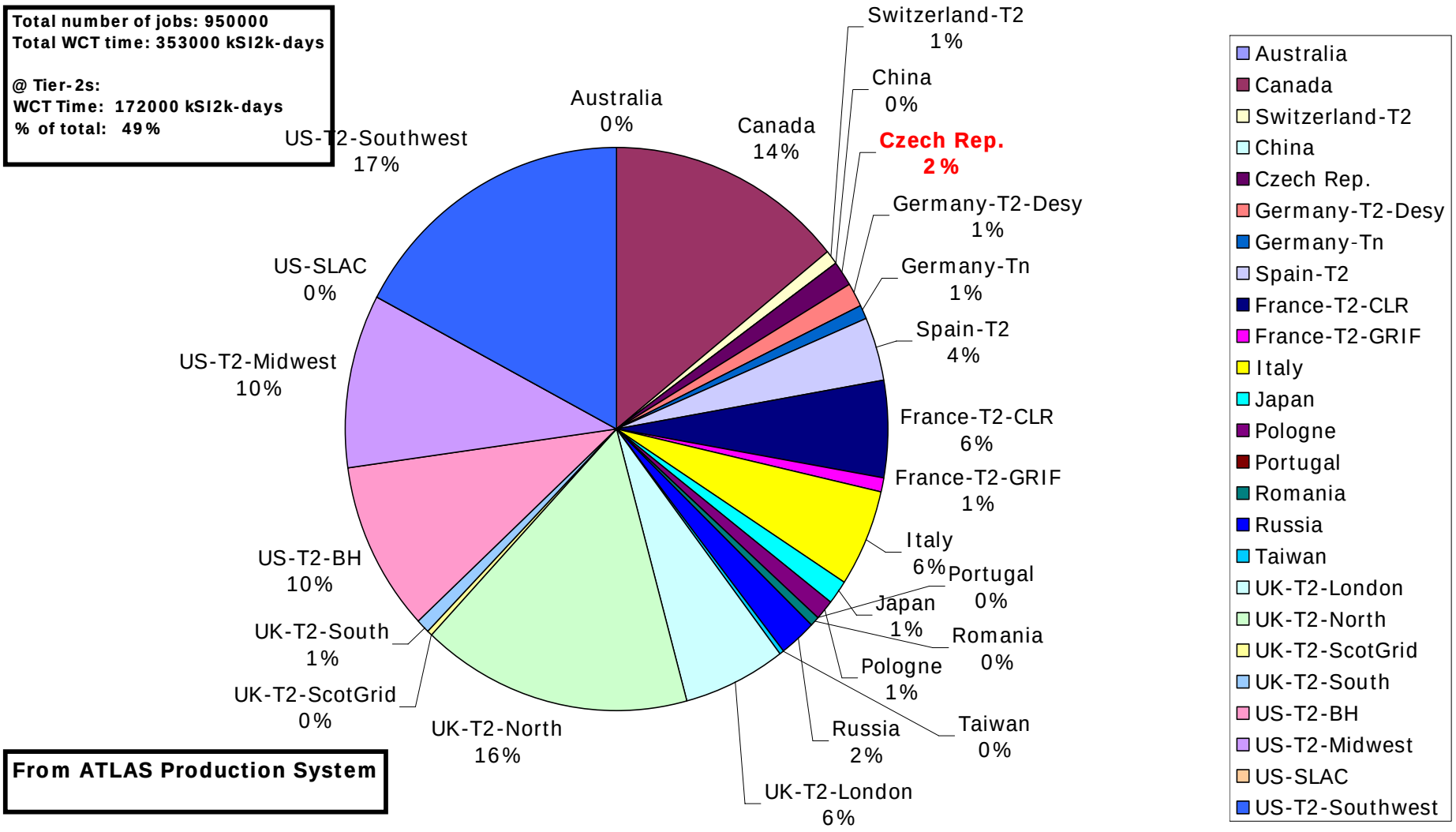




ALICE - production in 2006

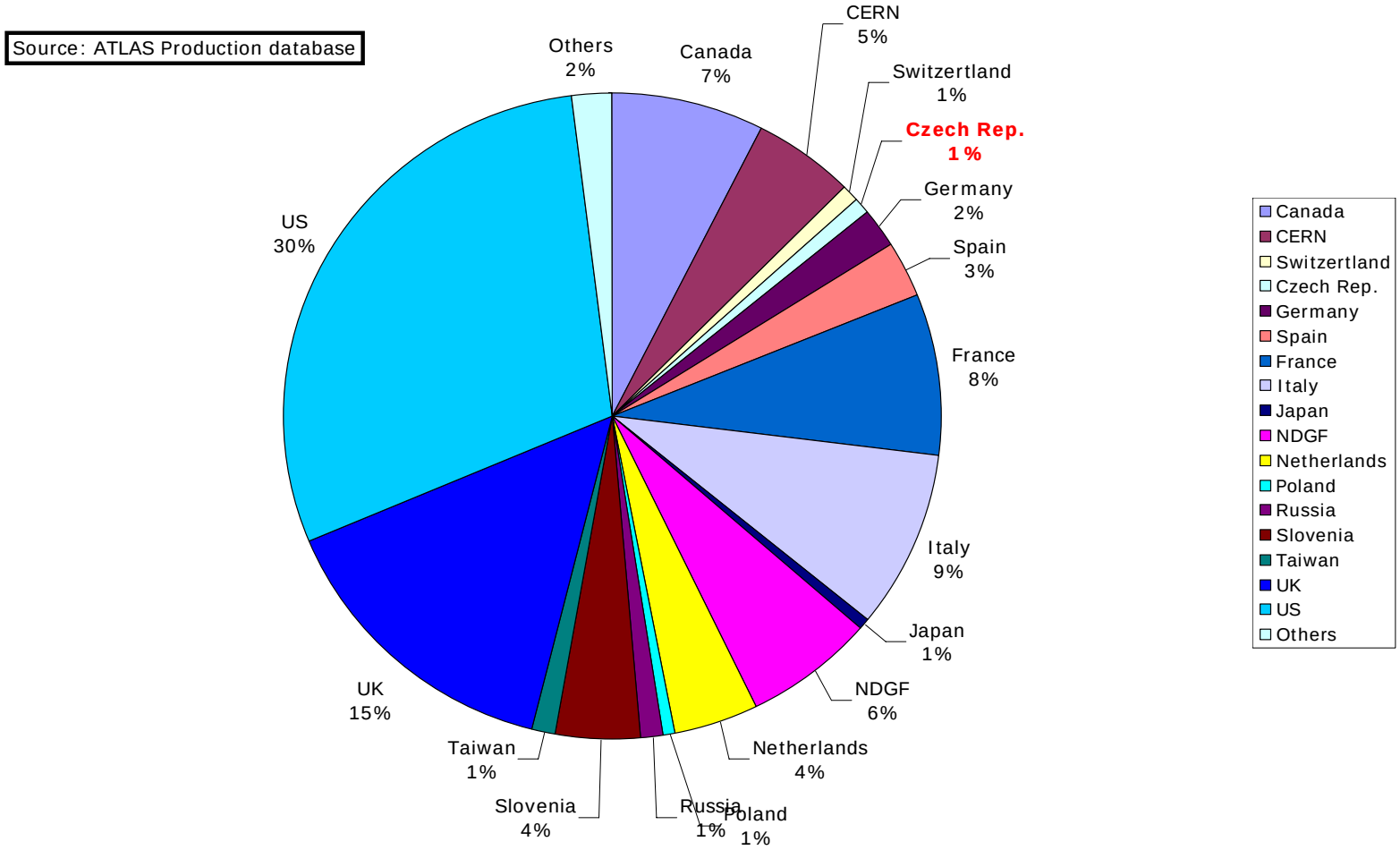


ATLAS Production at Tier-2s (Jan-Oct 2006) (Wall Clock Time)



Provided by Gilbert Poulard

ATLAS Production (Jan-Oct 2006) Wall Clock Time

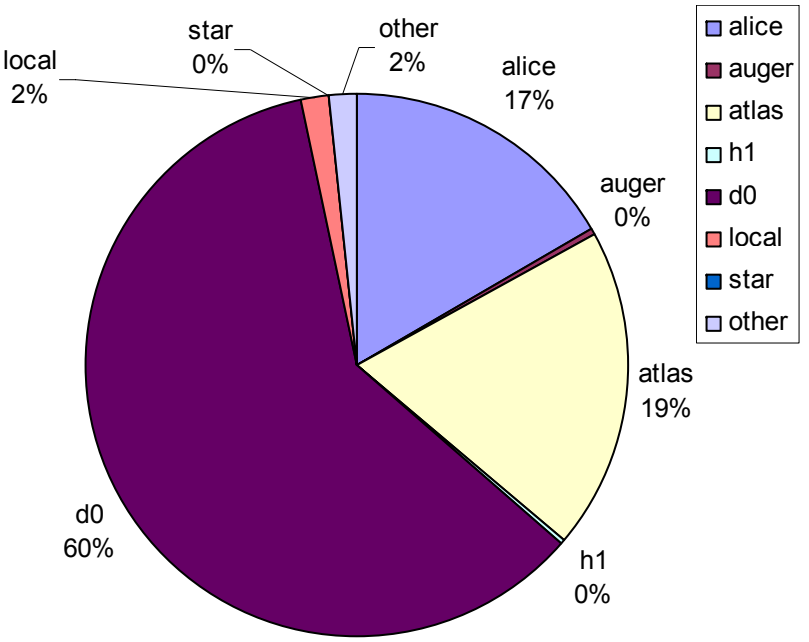


Provided by Gilbert Poulard

Statistics from PBS, GOLIAS

- **alice: 67601 jobs, 7975 days = 21 years**
- **atlas: 46510 jobs, 8931 days = 24 years**
- auger: 282 jobs, 146 days
- **d0: 132634 jobs, 28100 days = 76 years**
- h1: 515 jobs, 140 days
- star: 4806 jobs, 41 days
- long queue (local users): 676 jobs, 613 days

- **total: 262704 jobs, 46164 days = 126 years**



Fair-share ratios:

D0	ATLAS	ALICE+STAR	Others
50	19	7	23

These ratios were set at December when new machines arrived

Service planning for WLCG

- Proposed plan for WLCG Grid resources in the Czech Republic

	2007	2008	2009	2010	2011
ATLAS + ALICE					
CPU (kSI2000)	80	900	2000	3 600	5 400
Disk (TB)	10	400	1 000	2 160	3 400
Nominal WAN (Gb/s)	1	10	10	10	10

- Table based on WLCG updated TDR for ATLAS and Alice and our anticipated share in the LHC experiments
  - no financing for LHC resources in 2007 (current farm capacity share for LHC)
- We submit project proposals to various grant systems in the Czech Republic
- Prepare bigger project proposal for CZ GRID together with CESNET
  - For the LHC needs
  - In 2010 add 3x more capacity for Czech non-HEP scientists, financed from state resources and structural funds of EU
- All proposals include new personnel (up to 10 new persons)
- No resources for LHC computing, yet

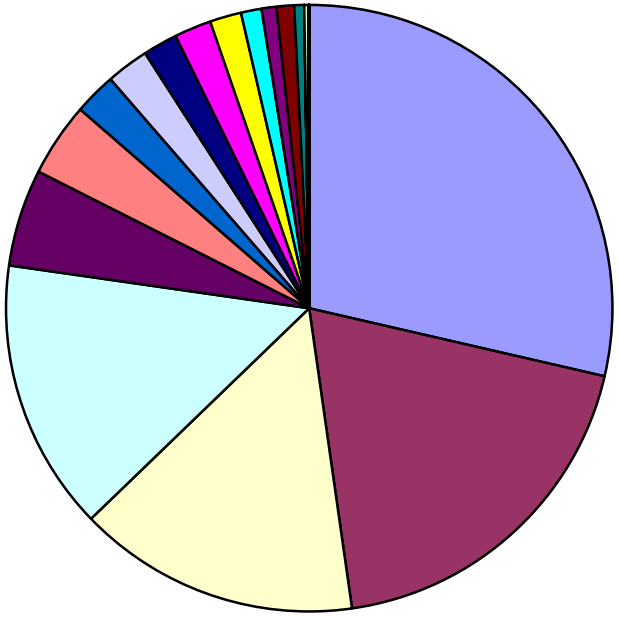
Storage - DPM

- 1 Head node (golias100) + 1 pool node (se2)
  - 1 more pool node in tests
- Mostly ATLAS files
  - 71608 files, 5.5 TB
- 2 pools, 5 filesystems:

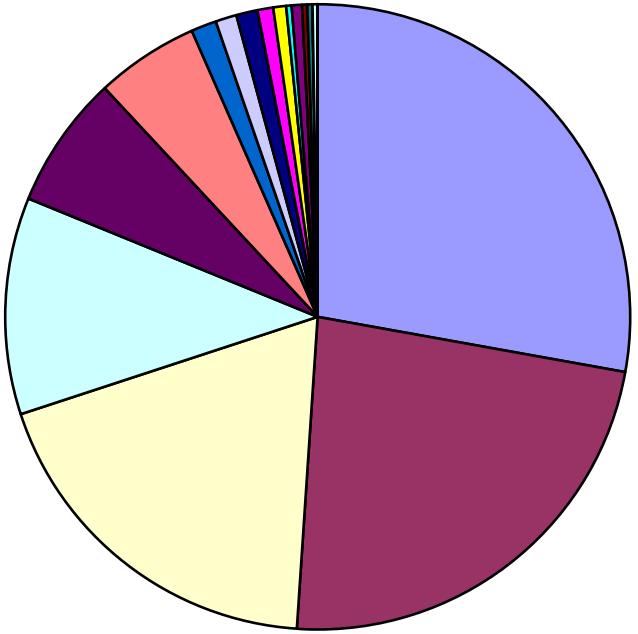
fs	N_FILES	
/mnt/array1	26131	1.8 TB, se2
/mnt/array1/	21804	1.9 TB
/mnt/array2/	11829	.9 TB
/mnt/array3	1075	.6 TB
/nbd_1	10769	1.0 TB, Network Block Device

Files per owner, size per owner

N_FILES per user



Space per user

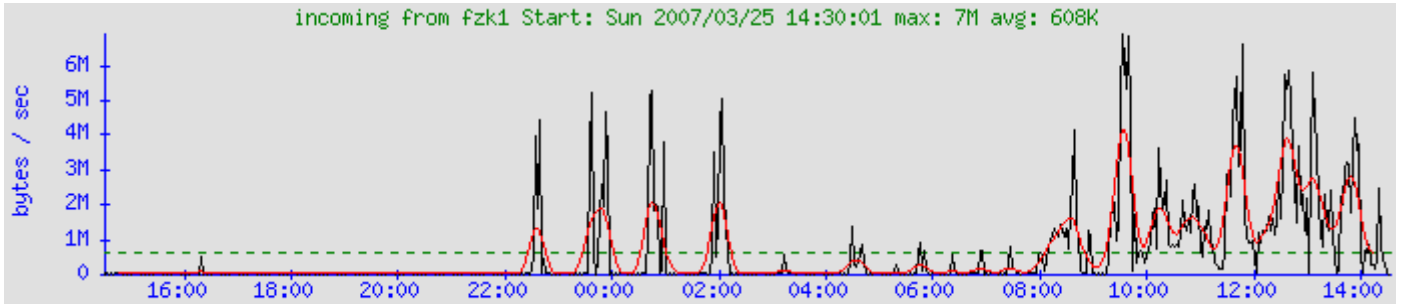


more than 90% used by production
(first nine most intensive users)
25 ATLAS users

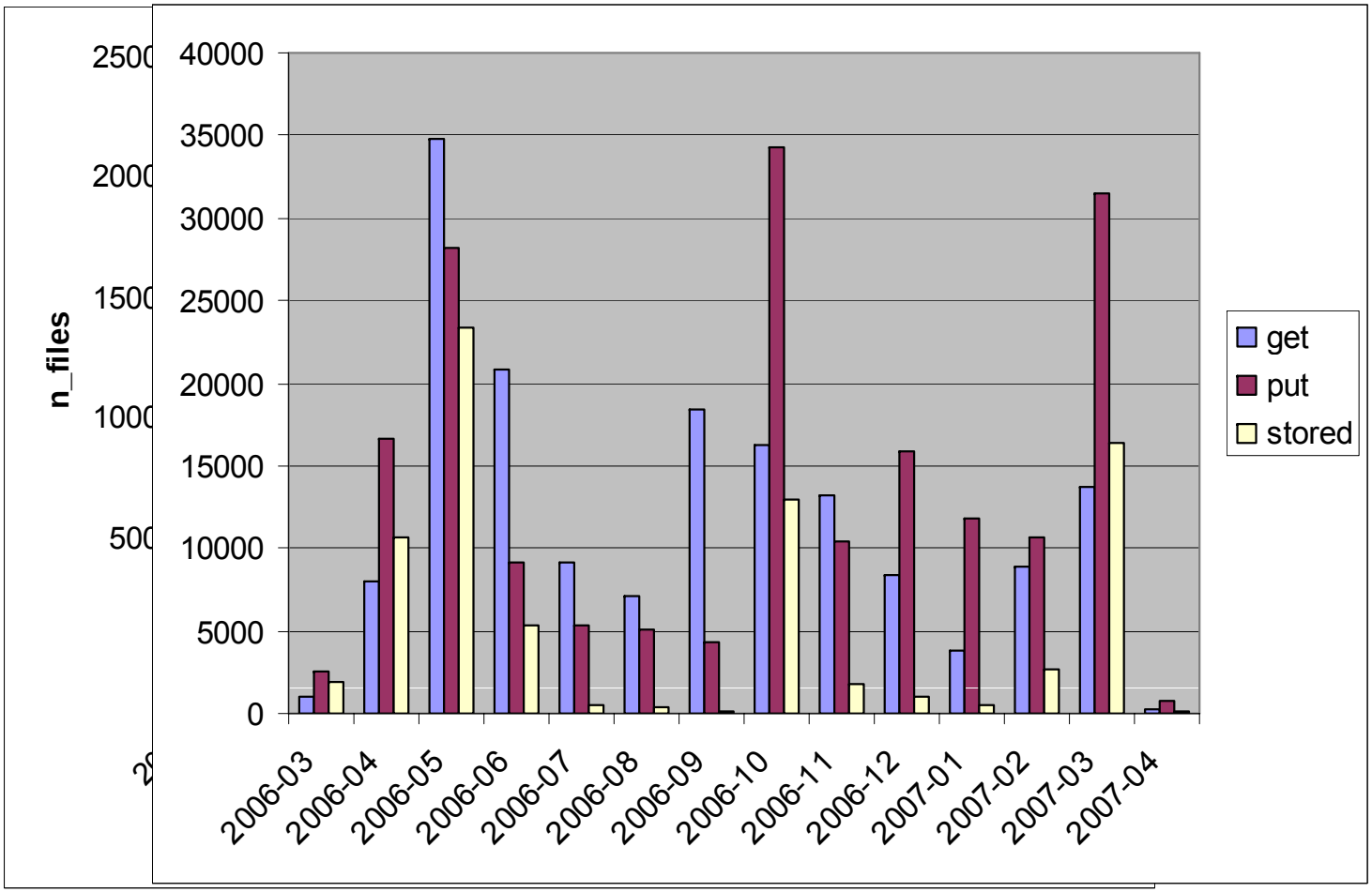
Access rates

- In total since 20.3.2006
  - 37 TB exported; 11 TB imported
  - NB: export average 100 GB/day
- using FTS: maximum 50 MB/s achieved for several hours
- import typically less than 5 MB/s from T1

(using DQ2 subscriptions)



DPM statistics



DPM statistics (cont)

All these graph hand made (SQL queries)

Shall we create scripts for monitoring?

Or will it be provided by middleware?

ACL in DPM

- all atlas groups are mapped to the same gid:

rowid	gid	groupname
1	1307	atlassgm
3	1307	atlas
7	1307	atlas/Role=lcgadmin
8	1307	atlasprd
10	1307	atlas/Role=production

- simplifies current setup
- change when secondary groups supported

Job priorities

- PBSPro, fair-share based on queues
- BDII, queue for users:

```
# golias25.farm.particle.cz:2119/jobmanager-lcgpbs-lcgatlas, praguelcg2, grid
    dn: GlueCEUniqueID=golias25.farm.particle.cz:2119/jobmanager-lcgpbs-lcgatlas,m
     ds-vo-name=praguelcg2,o=grid
    GlueCEUniqueID: golias25.farm.particle.cz:2119/jobmanager-lcgpbs-lcgatlas
    GlueCEStateEstimatedResponseTime: 5928394
    GlueCEStateFreeCPUs: 26
    GlueCEStateRunningJobs: 70
    GlueCEStateStatus: Production
    GlueCEStateTotalJobs: 83
    GlueCEStateWaitingJobs: 13
    GlueCEStateWorstResponseTime: 11856788
    GlueCEStateFreeJobSlots: 0
    GlueCEPolicyMaxCPUTime: 0
    GlueCEPolicyMaxRunningJobs: 70
    GlueCEPolicyMaxTotalJobs: 0
    GlueCEPolicyMaxWallClockTime: 0
    GlueCEPolicyPriority: 150
    GlueCEPolicyAssignedJobSlots: 0
    GlueCEAccessControlBaseRule: VO:atlas
```

Job priorities

- Queue for production:

golias25.farm.particle.cz:2119/jobmanager-lcgpbs-lcgatlasprod, praguelcg2, grid
dn: GlueCEUniqueID=golias25.farm.particle.cz:2119/jobmanager-lcgpbs-lcgatlaspr
od,mds-vo-name=praguelcg2,o=grid
GlueCEInfoContactString: golias25.farm.particle.cz:2119/jobmanager-lcgpbs-lcgatlasprod
GlueCEStateTotalJobs: 0
GlueCEStateWaitingJobs: 0
GlueCEStateWorstResponseTime: 0
GlueCEStateFreeJobSlots: 0
GlueCEPolicyMaxCPUTime: 0
GlueCEPolicyMaxRunningJobs: 0
GlueCEPolicyMaxTotalJobs: 0
GlueCEPolicyMaxWallClockTime: 0
GlueCEPolicyPriority: 150
GlueCEPolicyAssignedJobSlots: 0
GlueCEAccessControlBaseRule: VOMS:/atlas/Role=production

Problems

- Only gLite WMS understands

GlueCEAccessControlBaseRule: VOMS:/atlas/Role=production

Invisible for LCG RB

- Production can submit to the queue for users

Conclusion

- GRID for HEP in the Czech Republic
  - Has good grounds to be built-up
    - Excellent networking
    - Existing infrastructure of specialists contributing to international GRID projects
    - Pilot installation exists and is fully integrated into WLCG/EGEE and can be enlarged
  - Projects proposals for
    - HEP GRID - regularly proposing
    - National GRID proposal (with CESNET) prepared
  - No adequate financial resources for LHC computing, yet