ATLAS activities

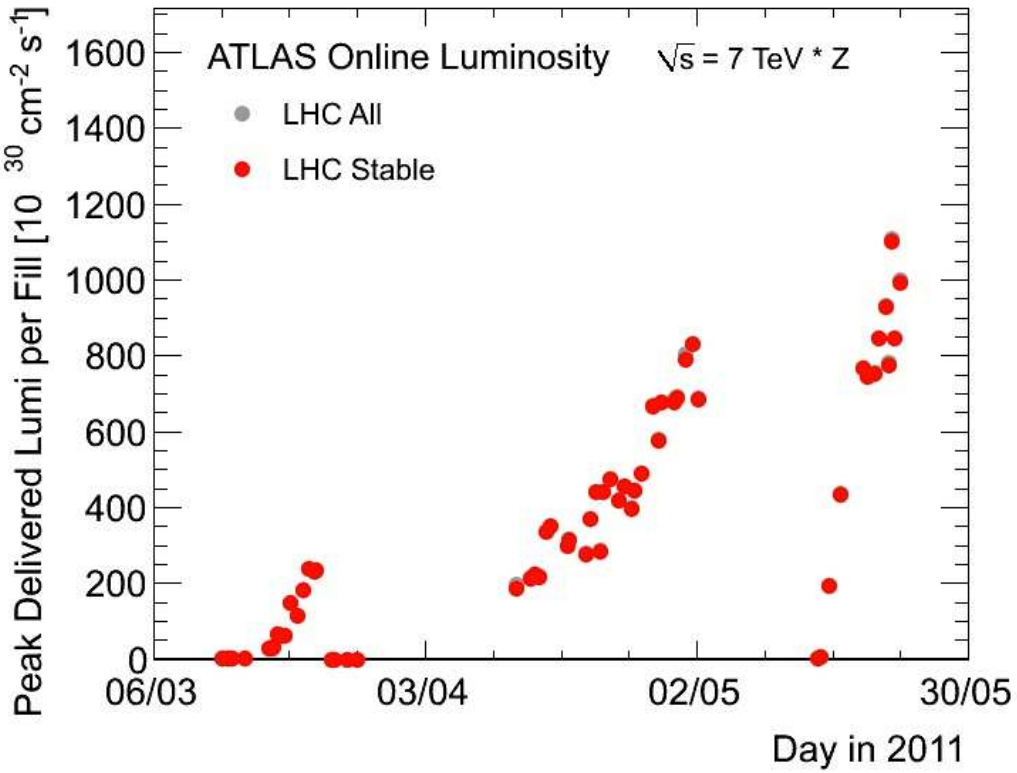
L. Poggioli, LAL Orsay

Thanks to many CAF colleagues
Special thanks to Eric Lançon

- Computing Model
- Resources
- Activities
- CAF
- FR-cloud

Basics

10^{33}!!!



Roadmap

- 5x10^{33} in a few months
- 20-25 pile-up events
- Run at higher rate
  - 400 Hz even 600 Hz
- Run in 2012

Impact on

- More load on T1 (reprocessing)
- More load on T2s (Simulation & analysis)
- Resources
  - CPU & Disk (more & bigger events)

Computing Model (1)

- Starting point
  - Rigid structure in clouds
  - Data pre-placed a priori
  - Data dispatched in many tokens (MC, data, physics groups)
  - Most datasets not used!
- Now
  - Much less pre-placed data
  - Data allocated dynamically via PD2P mechanism
    - Based on needs & popularity
  - Clouds inter-connectivity via T2D
    - Allows x-production across Clouds

Computing Model (2)

- 1 Copy of RAW data on DISK distributed over all T1s
  - RAW event is 1.4 MB (compression not viable yet)
- A rolling buffer of 10% of ESD distributed over T1s
  - ESD size reduced to 1.1 MB/evt
- "Small streams" ESD (~ 10% ESD) to T1s
- 10 copies of AOD distributed over all clouds
  - Only 2 copies of previous version AOD
- 10 copies of DESD distributed as AOD
  - Sum of all DESD = size of AOD

Everything OK?

- Running now smoothly with average μ up to 8
- Tier0 CPU time ~15s
  - OK for 400Hz with 50% LHC efficiency
  - New mechanism in place to dynamically allocate CERN batch resource in case of clogging (~double Tier0 capacity for short amount of time)
- AOD size ~200kB (data)
- Now talking about 600Hz, & μ up to 25 soon
  - First look: CPU x2 & AOD size x2
  - -> New round of scrutiny and improvements

Tokens

- Merging of MCDISK & DATADISK tokens done
- GROUPDISK usage
  - Has increased by a factor ~3.5 over last 10 months
  - Heavy usage (SM, Top) wrt low usage (Muons, Jets,eγ)
  - -> small usage (space, access) for most tokens
- Re-deployment of GROUPDISK tokens just'ed
  - Reduction of # tokens and enlargement of popular areas
  - Proposal to merge GROUPDISK areas with DATADISK & keep GROUPDISK only at T1s discussed
- Summary
  - To be kept @T2Ds where strong potential user community
  - DDM popularity tools -> take local access into account
  - For some small sites -> pre-placed group DS -> analysis OK
  - Political & Funding Agencies implications
    - T2s get funded partly because they host a GROUPDISK token

T2D

- Idea
  - T2 directly connected to T1s
  - Allows cross-cloud production
- Metrics
  - Based on transfer of small/medium/big DS (2GB)
  - Ongoing Sonar tests
- New candidates evaluated every 3 months
  - Today Tokyo, LAL, LPNHE
  - Next Clermont, Beijing
- Interaction with Lyon
  - For dedicated FTS channel update

See Sabine's talk yesterday

Databases

- ConDB & Frontier
  - 6 T1s + CERN with Frontier (incl. Lyon)
  - Existing Frontier servers will be consolidated
  - # FrontierT1s could possibly be further reduced
    - -> capacity used for additional TAG DBs
- TAG DB
  - 3TB per reprocessing – Need 2 copies
  - Look for 2 T1s CNAF & SARA (w/o Frontier)
- LFC
  - Today 16 (CERN, 10T1s, 5 US T2s)
  - Wish to group them in 1 @ CERN

Computing Resources

- Goal: to integrate 400 Hz trigger rate in 2011 FIXED resources
- Evaluate impact of pile-up on datasets size increase
- Resources
  - T0 a priory OK (mechanism to x2 capacity)
  - Small increase in CPU for T1s
  - For T2
    - CPU: Simul/analysis to be balanced, uncertainty on user part
    - DISK: Net increase

Resources @ T1s

Tier-1 CPU (kHS06)	2010	2011	2012	2013
Re-processing	34	27	51	51
Simulation production	108	109	83	87
Simulation reconstruction	3	34	22	29
Group (+user) activities	33	52	85	99
Total	*178*	*222*	*241*	*266*

CPU
-2011 small increase
-2012 tough

Tier-1 Disk (RB)	2010	2011	2012	2013
Current RAW data	3.9	4.3	4.3	4.3
Real ESD+AOD+DPD data	6.5	7.0	4.9	4.9
Simulated RAW+ESD+AOD+DPD data	3.9	4.2	4.7	5.6
Calibration and alignment outputs	1.8	0.4	0.4	0.4
Group data	2.3	4.0	5.6	6.5
User data (scratch)	0.6	0.6	0.6	0.6
Cosmics	0.2	0.2	0.2	0.2
Processing and I/O buffers	3.0	3.7	4.3	4.4
Total	*22.2*	*24.4*	*25.1*	*26.9*

DISK
-2011 OK
-2012 OK

TAPE
- 2011 & 2012 with 5%
RRB 2010

Resources @ T2s

Tier-2 CPU (kHS06)	2010	2011	2012	2013
Simulation production	65	65	65	68
Group activities	38	52	72	77
User activities	123	172	288	316
Total	**226**	**289**	**426**	**461**

Tier-2 CPU (kHS06)	2010	2011	2012	2013
Simulation production	65	65	65	89
Group activities	38	49	52	62
User activities	123	164	178	254
Total	**226**	**279**	**295**	**405**

CPU
-2011 smalll increase
-2012 tough

Tier-2 Disk (PB)	2010	2011	2012	2013
Current RAW data	0.4	0.0	0.0	0.0
Real AOD+DPD data	11	21	30	30
Simulated RAW+ESD+AOD+DRD data	8	11	14	19
Calibration and alignment output	0.0	0.3	0.3	0.3
Group data	3	6	8	8
User data (scratch)	1	2	2	3
Processing buffers	1	1	1	1
Total	**24**	**40**	**56**	**61**

Tier-2 Disk (PB) RRB Oct., 2010	2010	2011	2012	2013
Current RAW data	0.4	0.3	0.3	0.3
Real AOD+DPD data	11	20	20	29
Simulated RAW+ESD+AOD+DRD data	8	11	15	20
Calibration and alignment output	0.0	0.3	0.3	0.3
Group data	3	5	5	6
User data (scratch)	1	2	2	3
Processing buffers	1	1	1	1
Total	**24**	**38**	**44**	**60**

DISK
-2011 OK
-2012 Tough

- Simul/ana balance to be tuned
- If data doubles, not Simulation (1st order)

Computing Resources summary

- 400 Hz seems feasible
  - T0 headroom is small (Castor bandwidth & CPU)
  - Simulation at T1s reduced - More load on T2s
- Increase in RAW&AOD size CPU wrt to pileup under control
- BUT as mentioned before
  - With running at 600Hz & μ up to 25
  - Impact on CPU & data size might be problematic
- 2012 could be a tough year

Financial Resources

- LHC planning
  - 2013 switched to 2012
  - The machine works better and better
- France
  - 2011 budget cut
  - T1: -6%
  - T2s: -6-8%
- Cut in other countries also
- For 2012
  - Actions taken already?
  - Actions to be taken?

Manpower

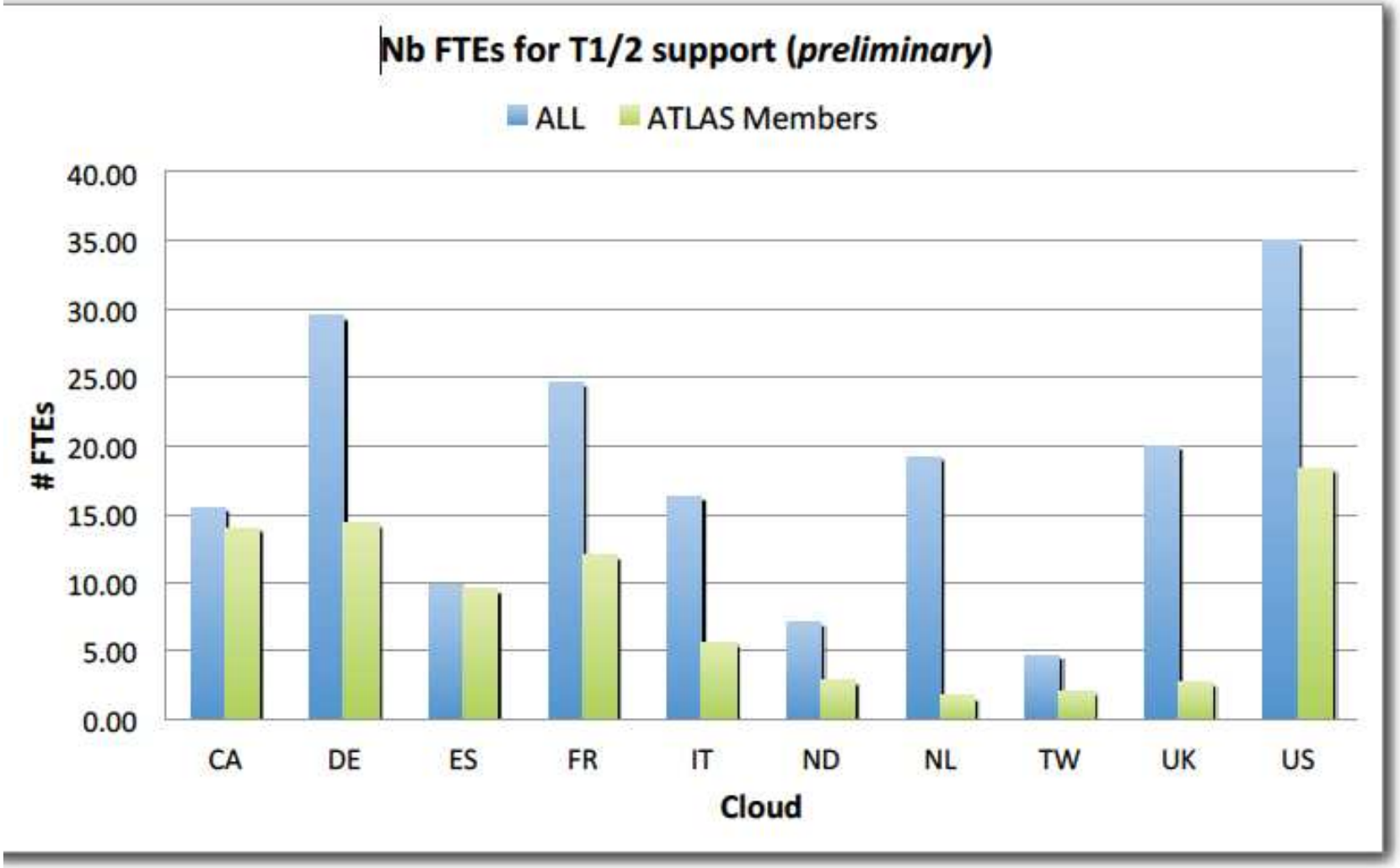
- Grouped in 3 classes
- Class 1
  - Handles data from detector-> T0 -> T1s OK
- Class 2 (remote shifts)
  - 15% deficit mostly in software validation
- Class 3 (FTE, Technical tasks)
  - 25% deficit in distributed computing development & sw infrastructure
- Handling direct support at T1/T2s

Support at T1/T2

- Undergoing study via Poll in ATLAS
  - Evaluate # persons providing ATLAS support at sites, and get proper credit (T1/T2/T3)
  - Via Class 4 OTP
- First preliminary results
  - 182 FTE (incl. 83 ATLAS FTE ie 155 ATLAS members). In total 430 (incl. 155 ATLAS members)
  - Similar size as class 3 (180 FTE)
  - For FR cloud: 7 (FR sites) + 4.3+4.7+1.3 (Japan+Romania+China)
- Under scrutiny in ATLAS management

FTEs per cloud

-



FTEs for FR

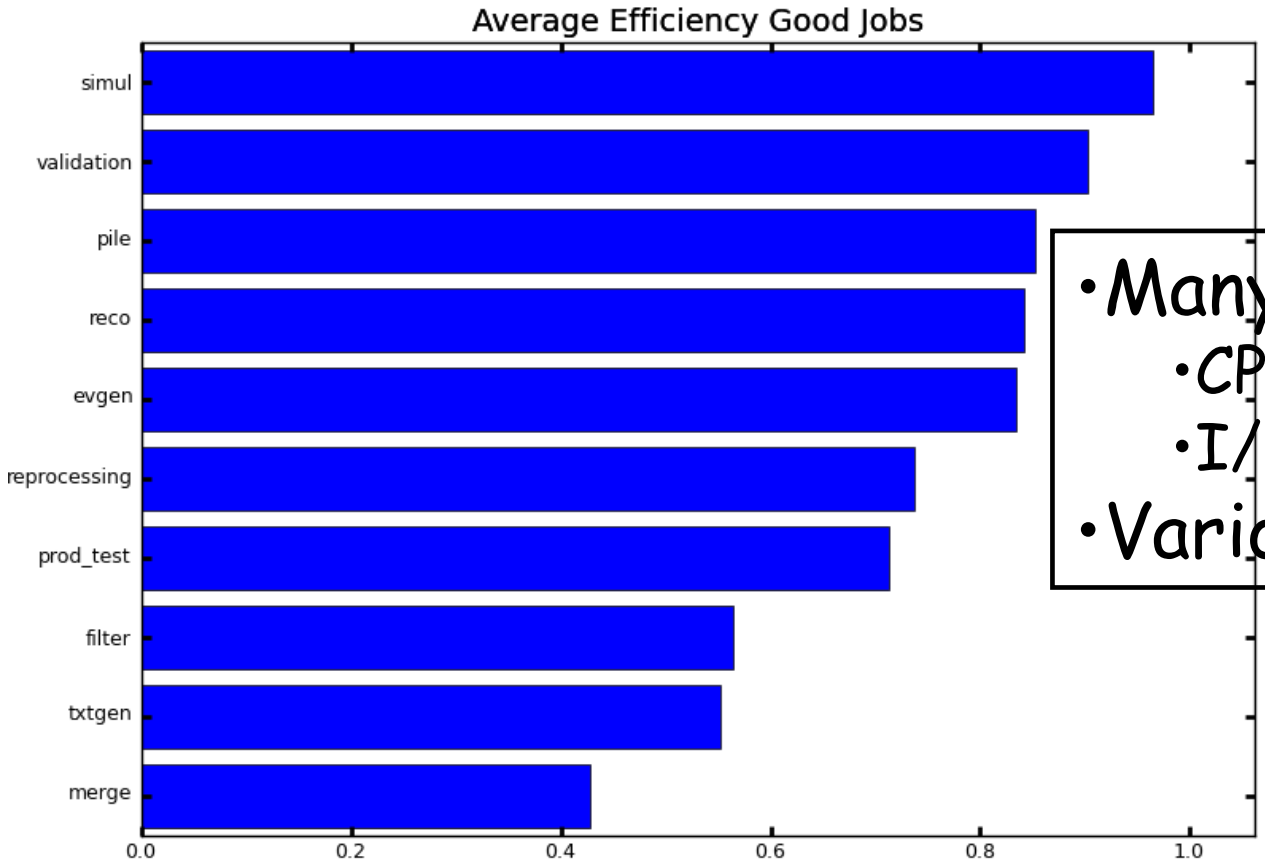
FR-CEA Tier-1	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	1.00
Saclay CEA	0.10	0.09	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	1.20
Annecy LAPP	0.12	0.11	0.12	0.11	0.12	0.11	0.12	0.12	0.11	0.12	0.11	0.12	1.37
Clermont - Ferrand	0.13	0.12	0.13	0.12	0.13	0.12	0.13	0.13	0.12	0.13	0.12	0.13	1.50
FR-CCIN2P3 Tier-1	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	1.00
FR-IN2P3 Tier-1	0.47	0.42	0.47	0.45	0.47	0.45	0.47	0.47	0.45	0.47	0.45	0.47	5.50
Grenoble LPSC													
LPNHE-Paris	0.07	0.06	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.80
Marseille CPPM	0.14	0.12	0.14	0.13	0.14	0.13	0.14	0.14	0.13	0.14	0.13	0.14	1.60
Orsay LAL	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.50
Chinese cluster	0.10	0.09	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	1.20
Tokyo ICEPP	0.37	0.33	0.37	0.35	0.37	0.35	0.37	0.37	0.35	0.37	0.35	0.37	4.30
Bucharest cluster	0.42	0.38	0.42	0.40	0.42	0.40	0.42	0.42	0.40	0.38	0.32	0.33	4.70

Lyon: 6.5 FTE + 1 FTE (direct support ATLAS)

FR-cloud T2s: 7 (FR sites) + 4.3+4.7+1.3 (Japan+Romania+China)

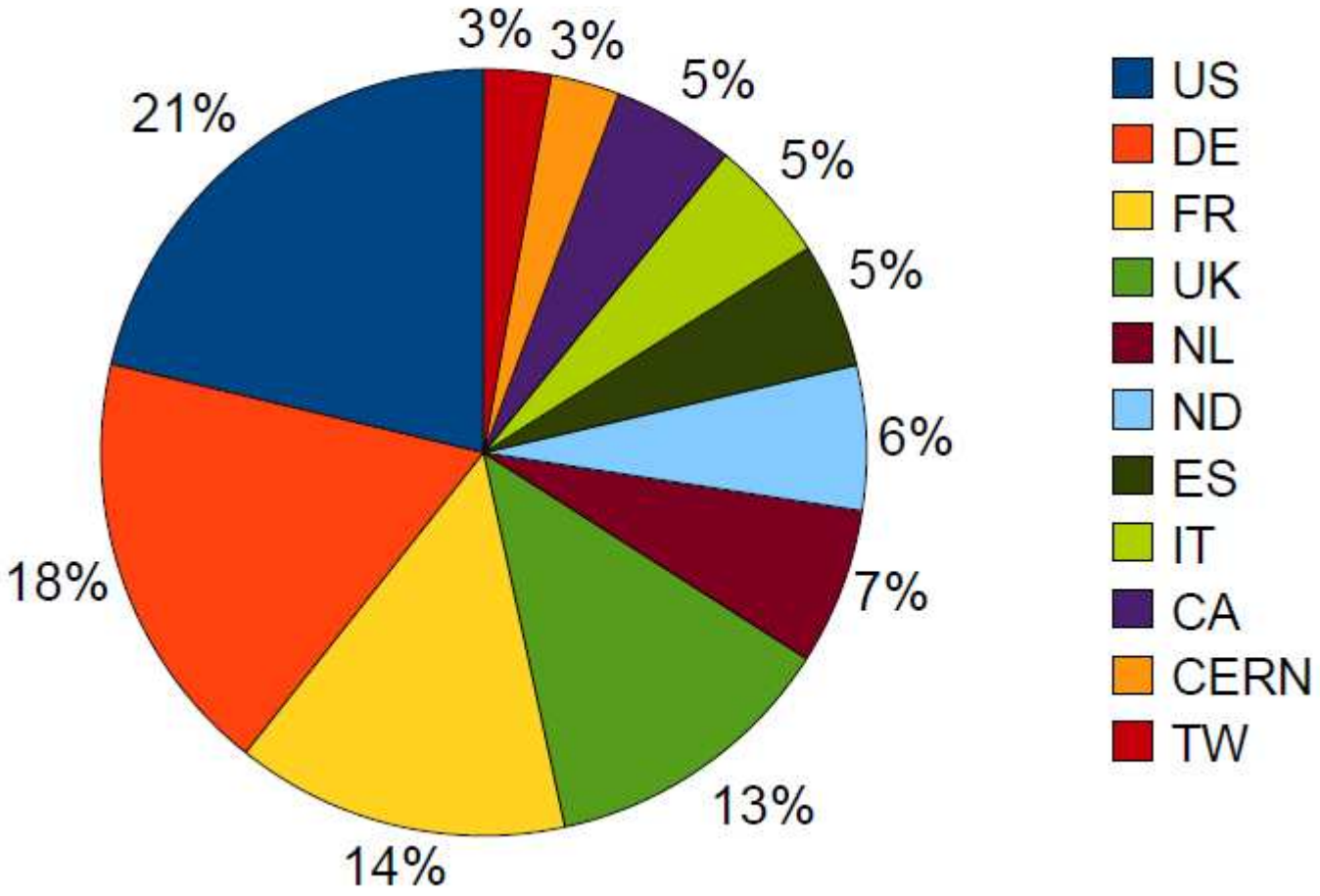
ATLAS activities

- Production (MC, mostly at T2)
- Reprocessing (Data & MC, at T1)
- Analysis (T2)

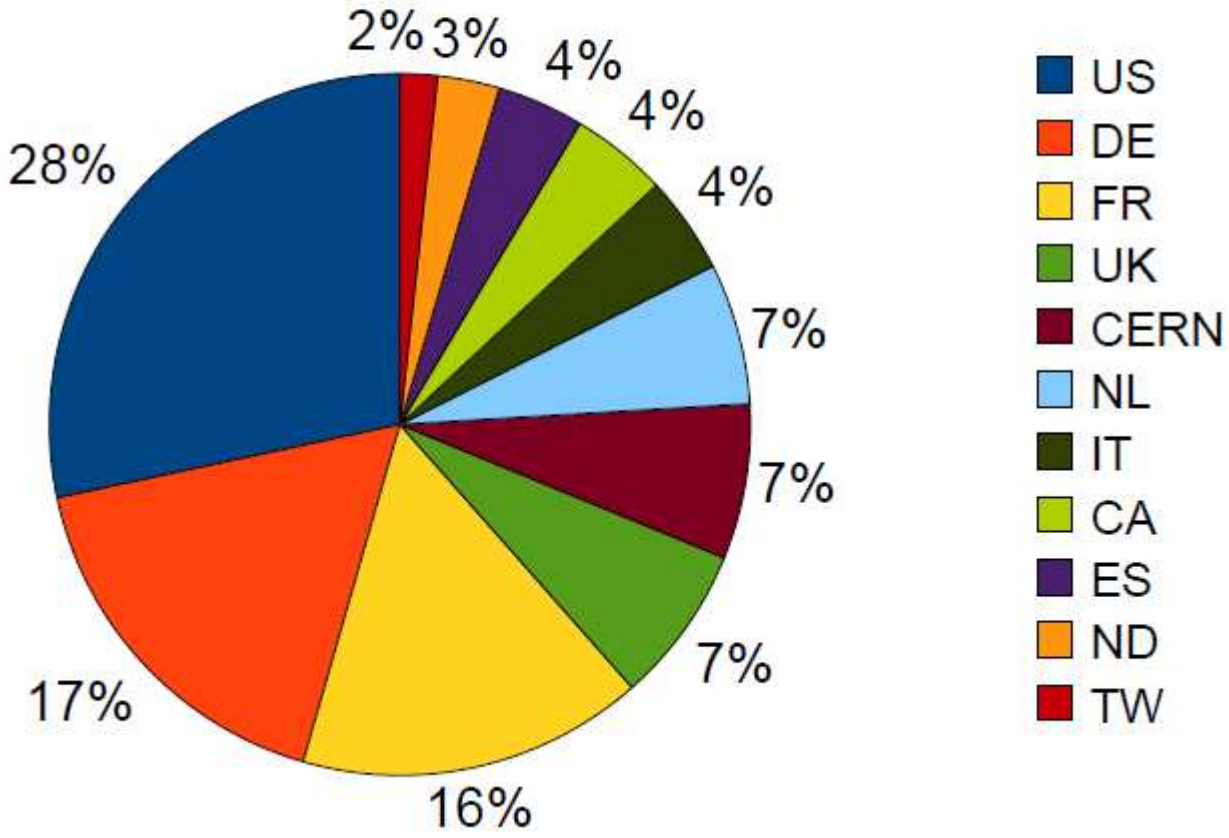


- Many different tasks
  - CPU usage
  - I/O
- Various efficiencies

Clouds activity : production March-May 2011

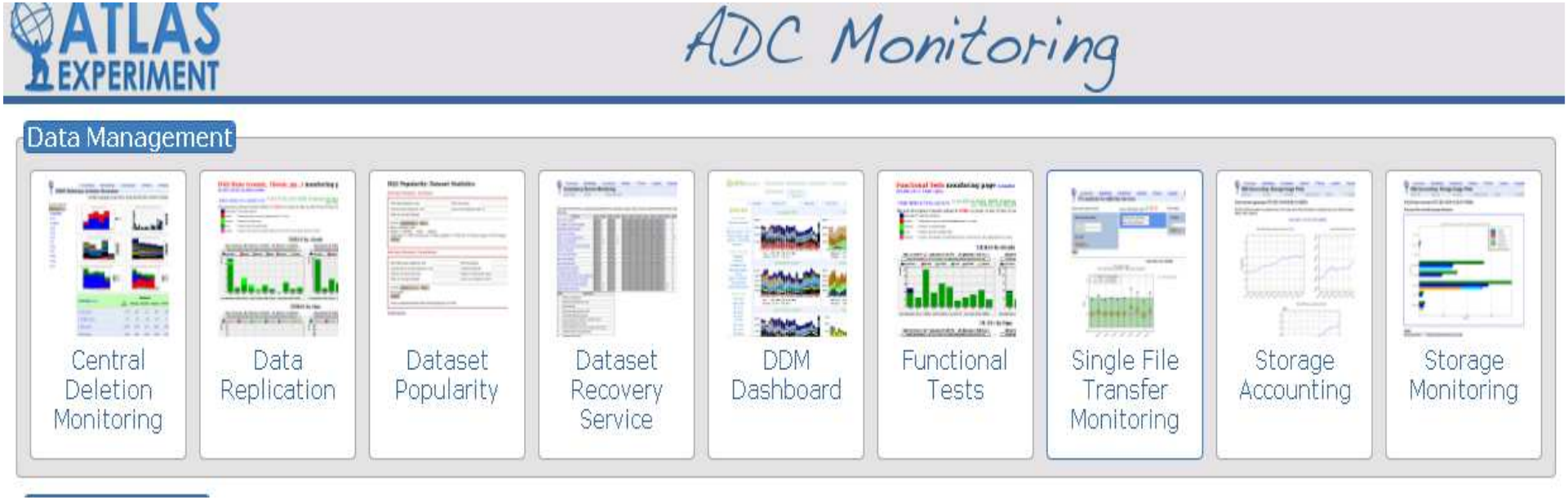


Clouds activity : analysis March-May 2011

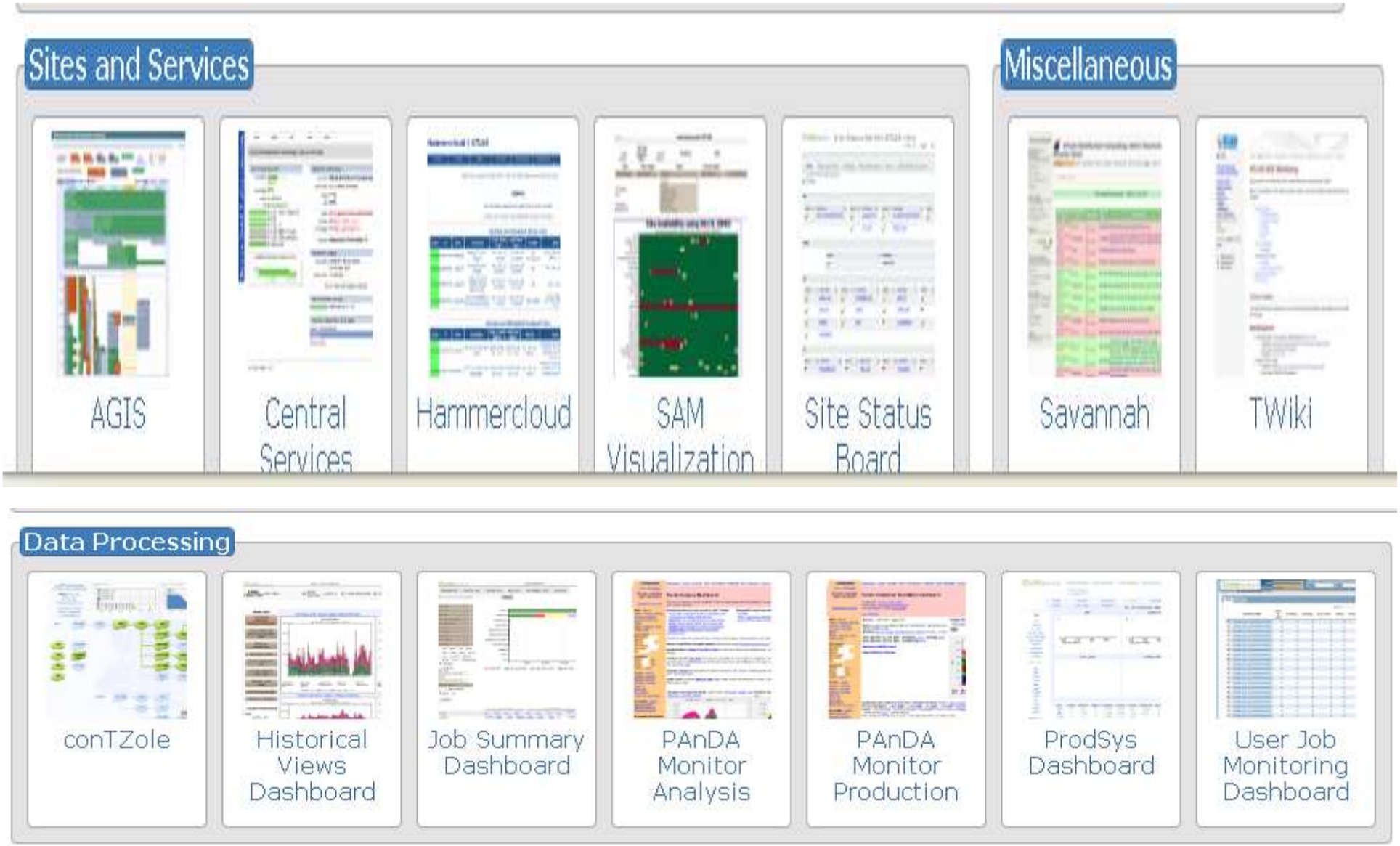


ATLAS Monitoring (1)

- Group all ATLAS monitoring tools in one page: http://adc-monitoring.cern.ch/
- 3 groups
  - -Data Management
  - -Data Processing
  - -Sites & Services



ATLAS Monitoring (2)



FR-SQUAD activities

- T3 share part implementation
- T2D implementation
- Network performance
- Factory/SchedConfigs/voboxes
- CREAM-CE
- Follow-up new batch system at Lyon
- Interplay, VL queues & T1/T2 at Lyon
- Monitoring improvement
- Overall communication
- Follow-up on Romanian sites
- CVMFS implementation
- Hammercloud tests

Sabine, Wenjing, Irena, Emmanuel, LP

Plus the standard FR-cloud operation daily follow-up/actions

CAF-PAF interaction

- Regular meeting with French physicists
  - Better understand Analysis problems & needs

Labo	Which Physics Group	Which data format?	Local T3 used for analysis?	Which resources needed at Lyon (interactive, batch, storage)
CPPM	Top, W/Z, Higgs, B-tagging, Leption ID, B-jet trigger	TopD3PD, SM D3PD-> NTUP (prun) HSG2 D3PD, AOD + prun DESDM_TRACK + prun Egamma D3PD B-jet trigger D3PD	Yes (local batch & storage , including proof)	Proof, FLAVTAG GROUPDISK token
LAL	Higgs, Egamma, SUSY	D3PD + prun ESD (Egamma)	Yes (LOCALGROUPDISK)	LOCALGROUPDISK /sps (4TB, PAU with quota)
LAPP	SM (WZ,γ) + Excotics	D3PD skim/slim -> NTUP	Yes if Grid space visible from local batch	LOCALGROUPDISK ~20TB
LPC	Top, Z', Jets	D3PD->Skimmed D3PD	Yes (local batch & storage)	No particular needs
LPNHE	Top, SUSY, Higgs	D3PD + ESD (prun)	Yes (LOCALGROUPDISK)	/sps (a bit)
LPSC	single-Top, Jets e, γ	NTUP common D3PD	No, use prun on the Grid	30TB LOCALGROUPDISK /sps + proof for SUSY
Saclay	SM, Top, Higgs	SM, Top D3PD (prun)	Yes (local batch & storage)	GROUPDISK (SM) LOCALGROUPDISK ~40TB

Handling GROUPDISKs

- Most tokens not used:filled

Site	Group	Role	Used by role(TB)	Booked by role(TB)	ΣUsed(TB)	ΣBooked(TB)
BEIJING-LCG2	SOFT-TEST	/atlas/soft-test/role=production	0	1.1	0.0	1.1
GRIF-IRFU	PHYS-TOP	/atlas/phys-top/role=production	53.21	54.98	53.21	54.98
GRIF-LAL	PHYS-SUSY	/atlas/phys-susy/role=production	17.33	27.49	17.33	27.49
GRIF-LPNHE	PHYS-SM	/atlas/phys-sm/role=production	23.79	27.49	23.79	27.49
IN2P3-CC	PERF-EGAMMA	/atlas/perf-egamma/role=production	35.62	54.98	35.62	54.98
	PERF-FLAVTAG	/atlas/perf-flavtag/role=production	21.93	27.49	21.93	27.49
	PERF-JETS	/atlas/perf-jets/role=production	28.45	54.98	28.45	54.98
	PERF-MUONS	/atlas/perf-muons/role=production	17.78	27.49	17.78	27.49
	PHYS-BEAUTY	/atlas/phys-beauty/role=production	0.22	27.49	0.22	27.49
	PHYS-HIGGS	/atlas/phys-higgs/role=production	12.88	27.49	12.88	27.49
IN2P3-CPPM	PERF-FLAVTAG	/atlas/perf-flavtag/role=production	0.08	5.5	0.08	5.5
IN2P3-LAPP	PHYS-SM	/atlas/phys-sm/role=production	29.0	54.98	29.0	54.98
IN2P3-LPC	PHYS-TOP	/atlas/phys-top/role=production	28.51	49.48	28.51	49.48

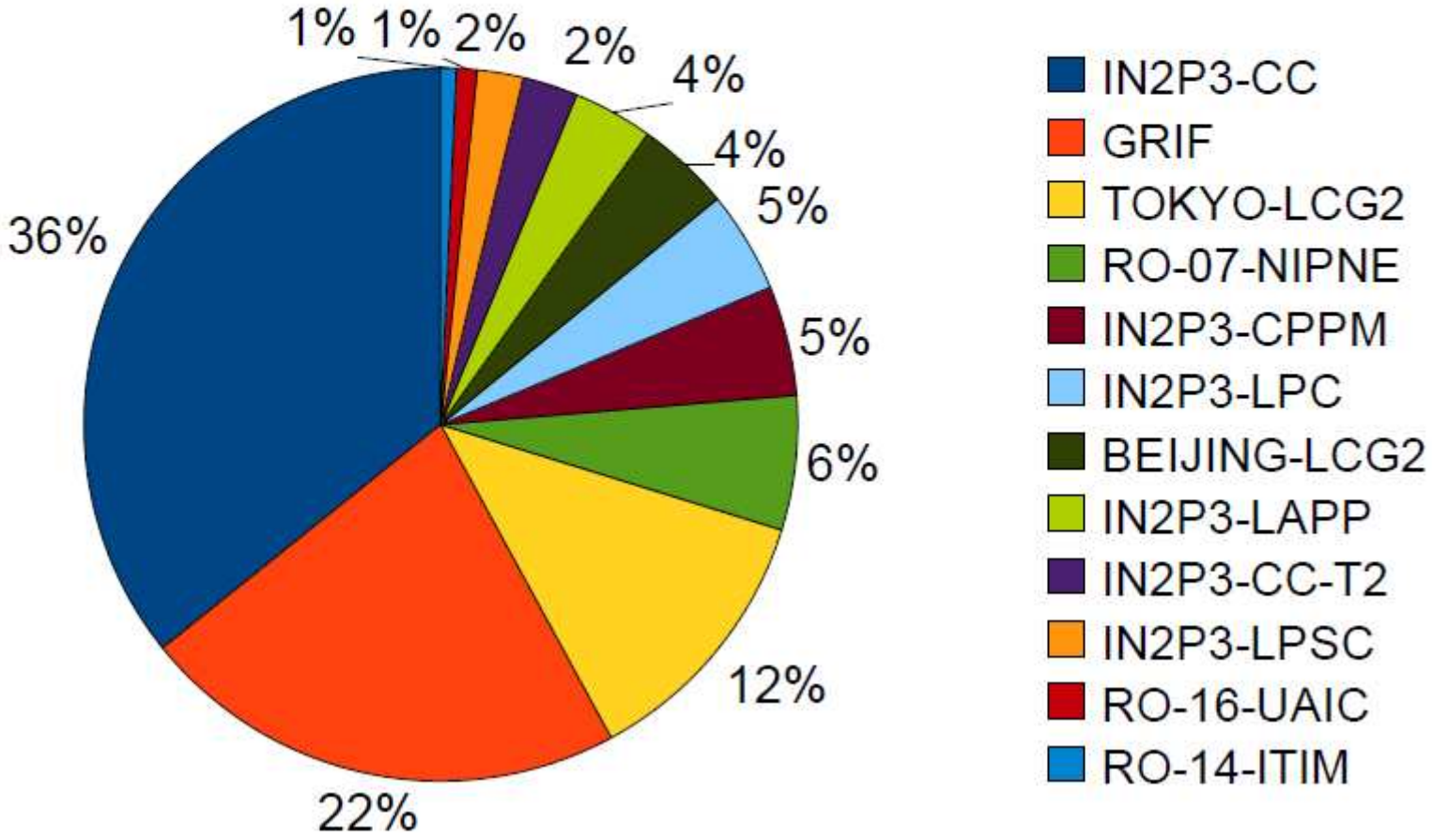
- For 2011 : group production stored at T1s and at T2Ds
- Storage at T1 : to be specified when task defined (FR physicists should ask for it)
- Which tokens needed on FR T2s?

See Emmanuel's talk this afternoon

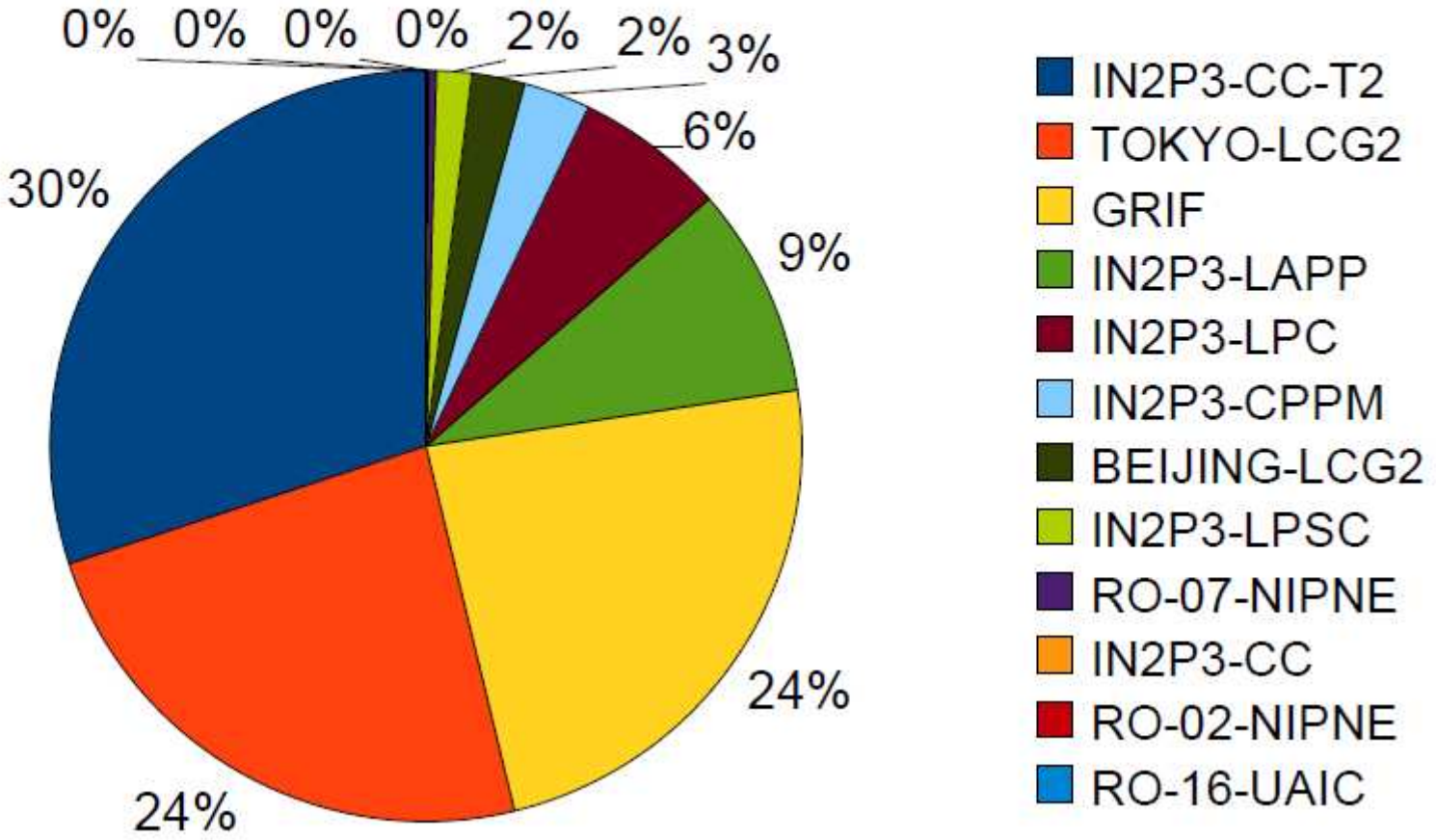
FR sites

- T2s/T3
  - No major problem
  - See reports in yesterday session
  - New: LPSC request T3 to T2
  - New: 4 Romanian sites fully operational!!
- Lyon
  - Monitoring
  - CVMFS
  - Reprocessing
  - Problems

Sites activity : production March-May 2011

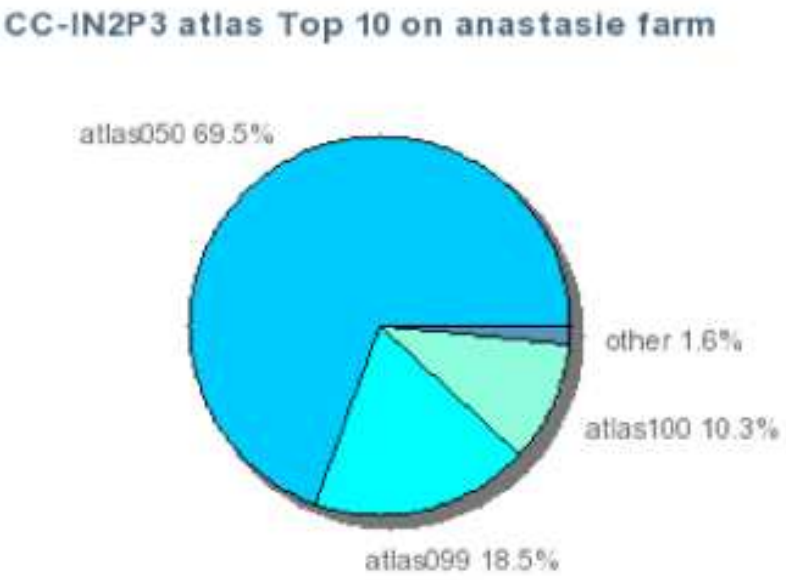


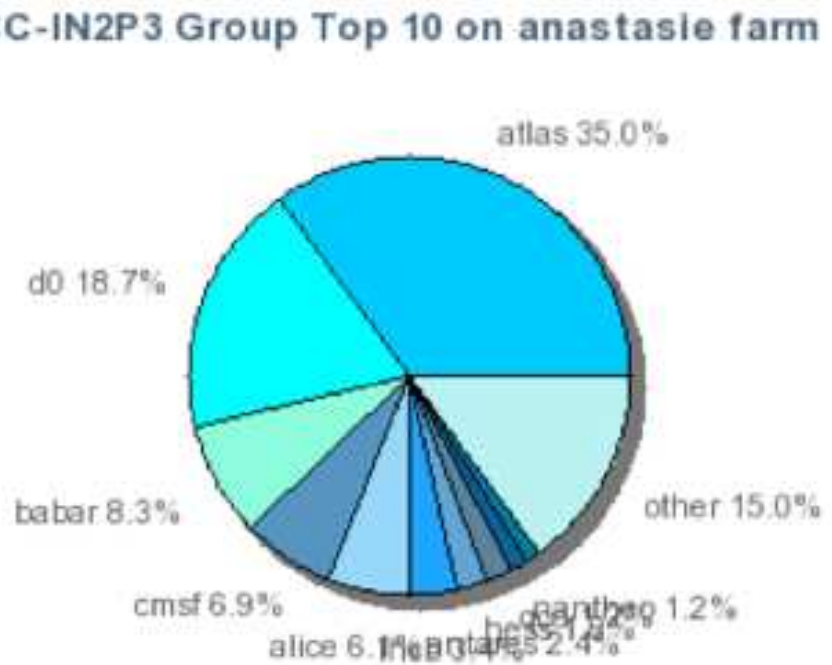
Sites activity : analysis March-May 2011



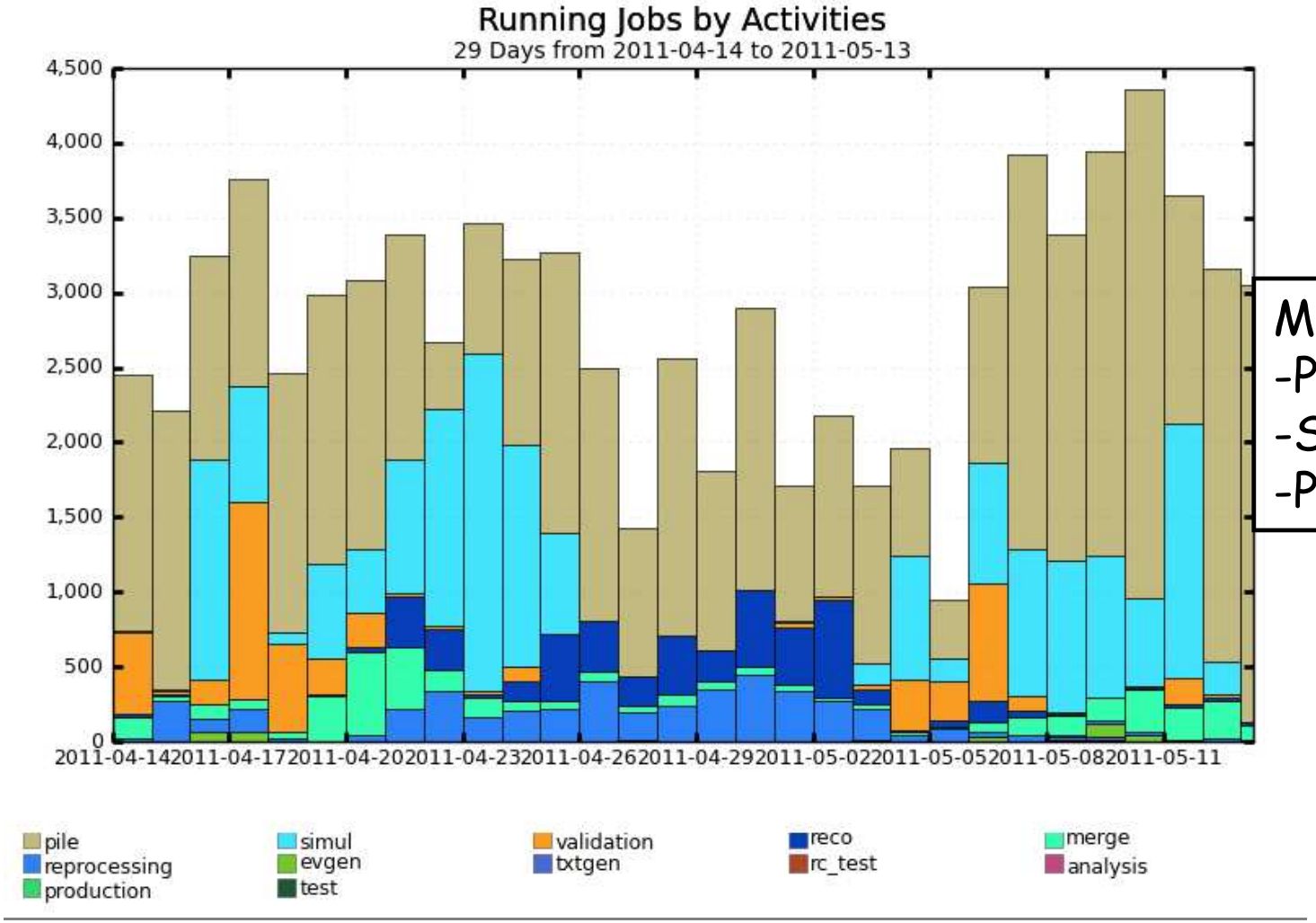
ATLAS@CC in 2010

- 35% of CC CPU consumption
- ATLAS
  - 70% T1
  - 19% Analysis
  - 10% T2 production
  - 2% Ganga/bqs



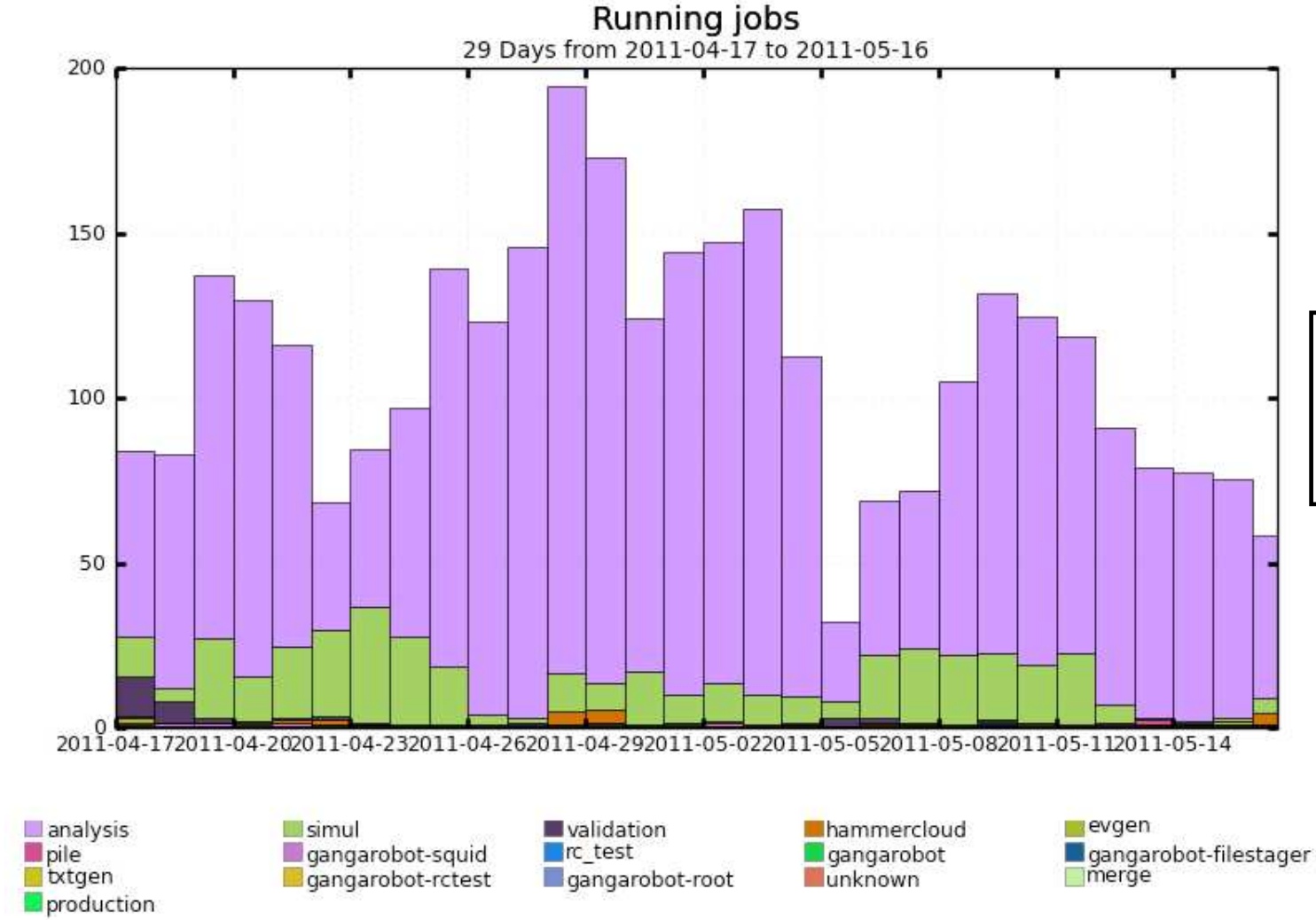


Activities @ Lyon T1



Mostly
-Pileup
-Simulation
-Period dependent

Activities @ Lyon T2



Mostly
-Analysis
-Simulation

Problems

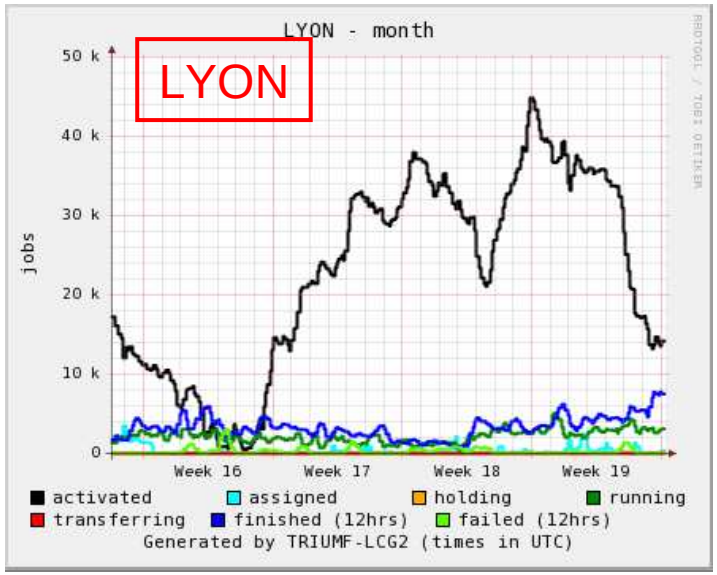
- Reprocessing
- Backlog of activated jobs
- Poor CPU efficiency
- /afs – SW releases

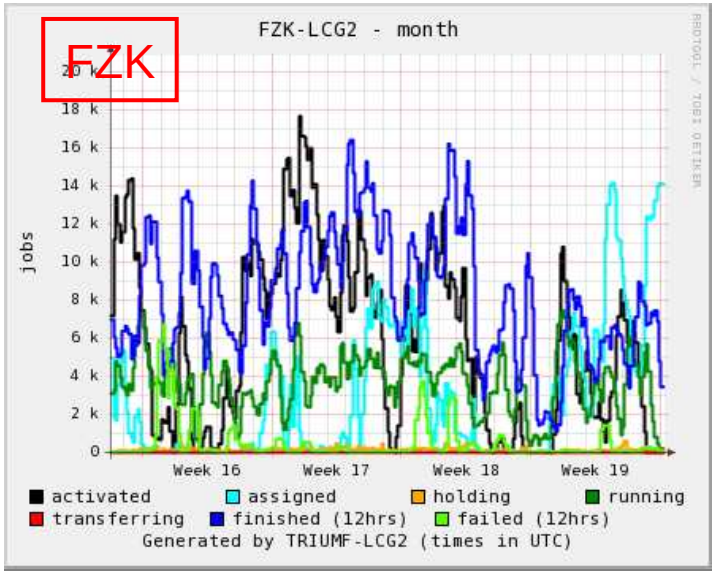
Reprocessing

- Major failure of November reprocessing
  - dCache issue
- Causes
  - New team
  - HW config problems
  - System unstable for long period
- Now Improvement with dcache experts
  - checksum calculation
  - load balancing improvement
  - network usage internal only

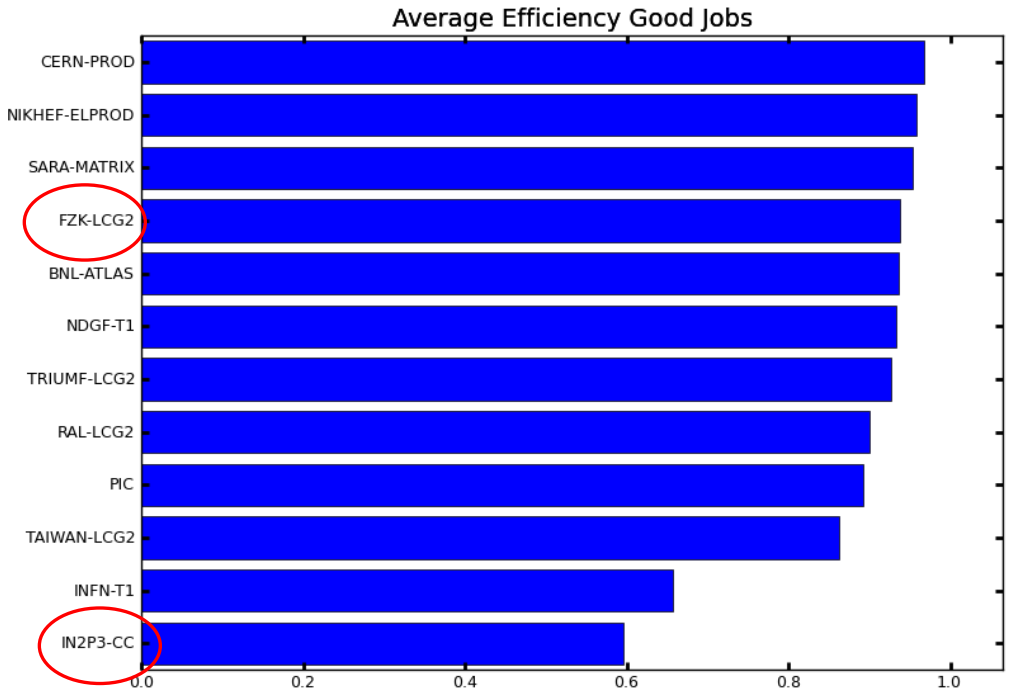
Backlog of activated jobs

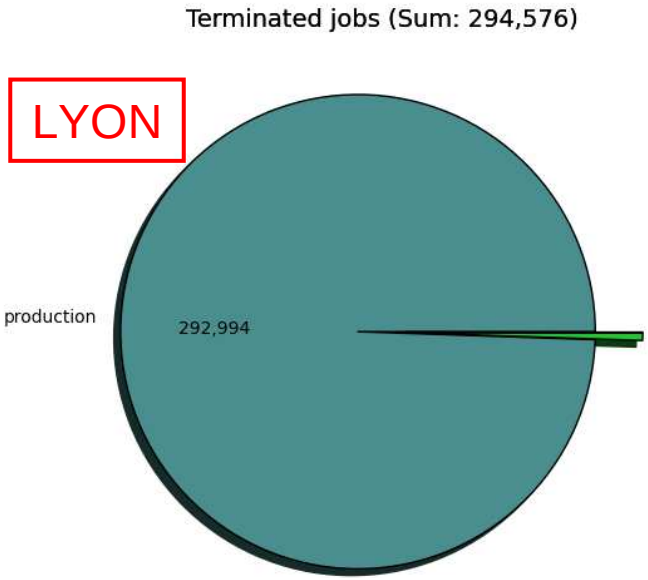
- Last month (up to 40k activated jobs!!!)
- Induces huge backlog in MC reprocessing
  - Transferred to T2s

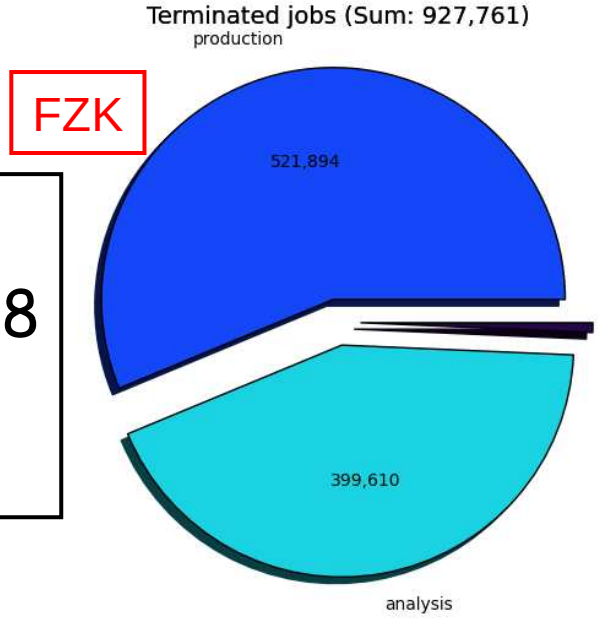




Efficiency problem



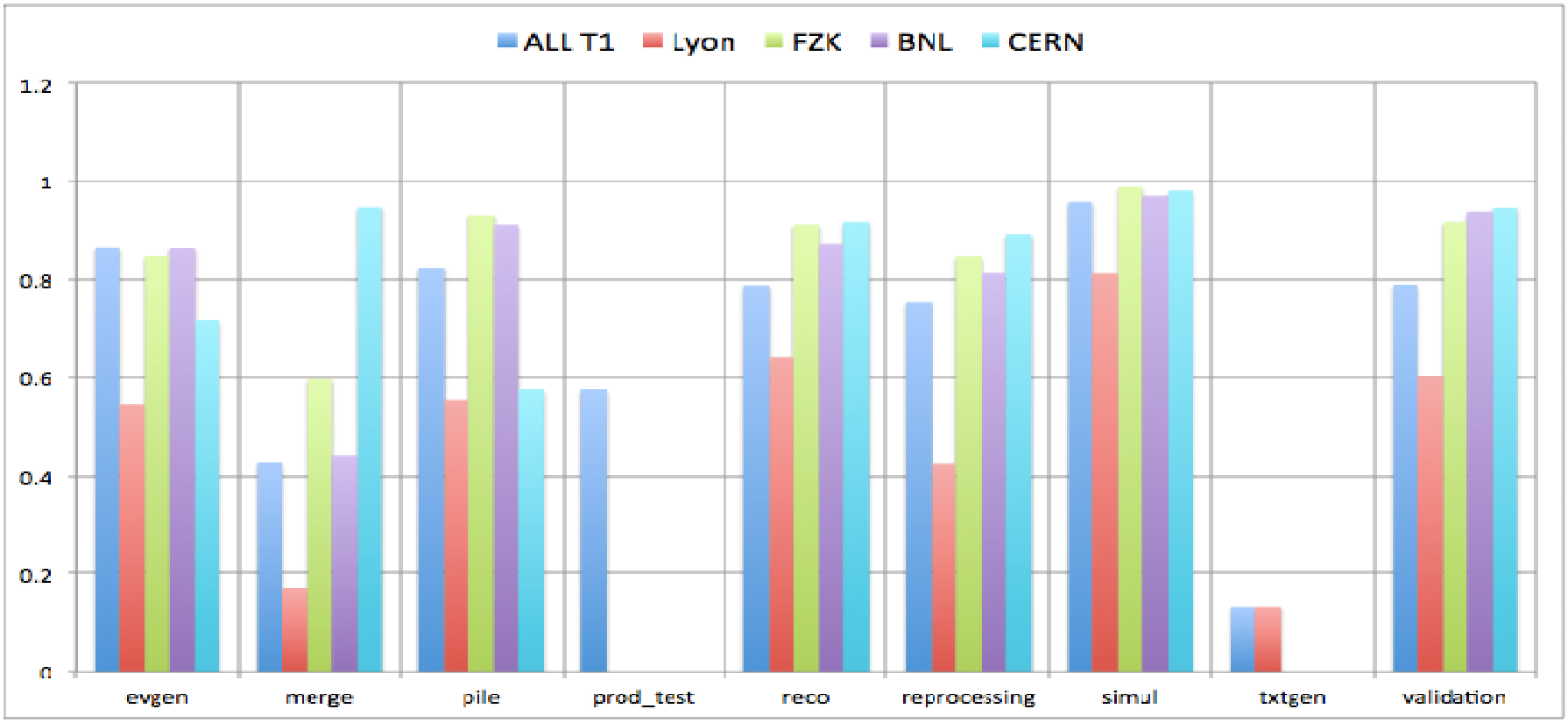




Over the last month

- #Production jobs (FZK/LYON) = 1.8
- An dFZK does 43% analysis

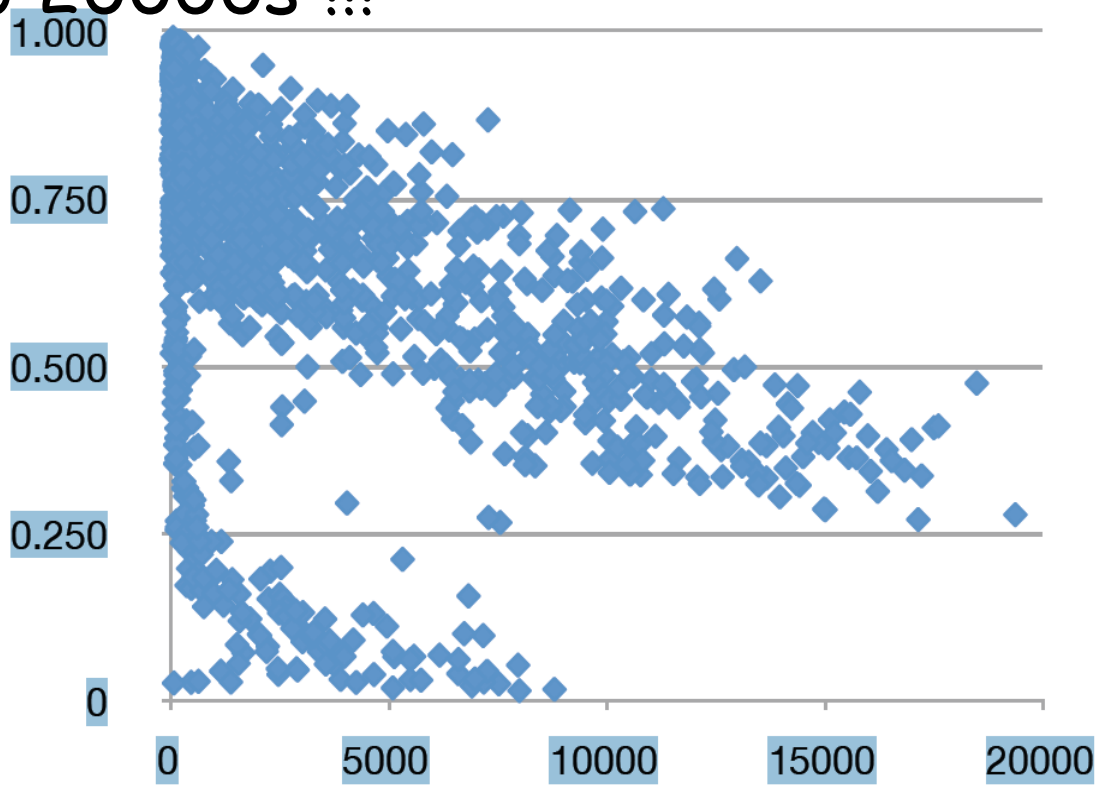
CPU efficiency by activities



- Simulation : Lyon below the average
- For heavy input load (merge, reprocessing)
  - Lyon is far below the average : storage ?

/afs issue

- CPU efficiency <-> AFS client cache saturation
  - when too many jobs run on same x-core machines
- Efficiency directly linked to setup times
  - Up to 20000s !!!



E. Lançon
This weekend

CernVM-FS

- Essential to have it @ CC
  - LHCb at many T1s
  - Also in ATLAS
    - T3s, LPNHE, Beijing soon
- Would avoid /afs problems
  - With setup time
  - With release installation
  - Support with stand install (deSalvo) will vanish

To be improved: Monitoring

- Interaction CAF/CC to better define needs
- Better monitoring of key component needed
  - FTS
  - dCache / SRM
  - Batch
- CPU monitoring :
  - Consumption T1, T2, T3 in form of CPUs, Not # jobs
    - Available : sum T1+T2+T3 CPU consumption vs time available
  - Compared to pledge values for T1 & T2
- FTS monitoring :
  - History and throughputs in graphical format

Would allow Pro-activity from CC

To be improved: Support

- Departure of C. Biscarat: huge loss
- Exceptional interlocutor: Pierre Girard
- But we need a more solid support
  - Better continuity in support
  - Better understanding of ATLAS sw
    - Ability to test efficiency wrt large variety of jobs
  - Allows pro-activity
  - Allows ATLAS to be kept informed