ALICE Operations short summary and directions in 2012

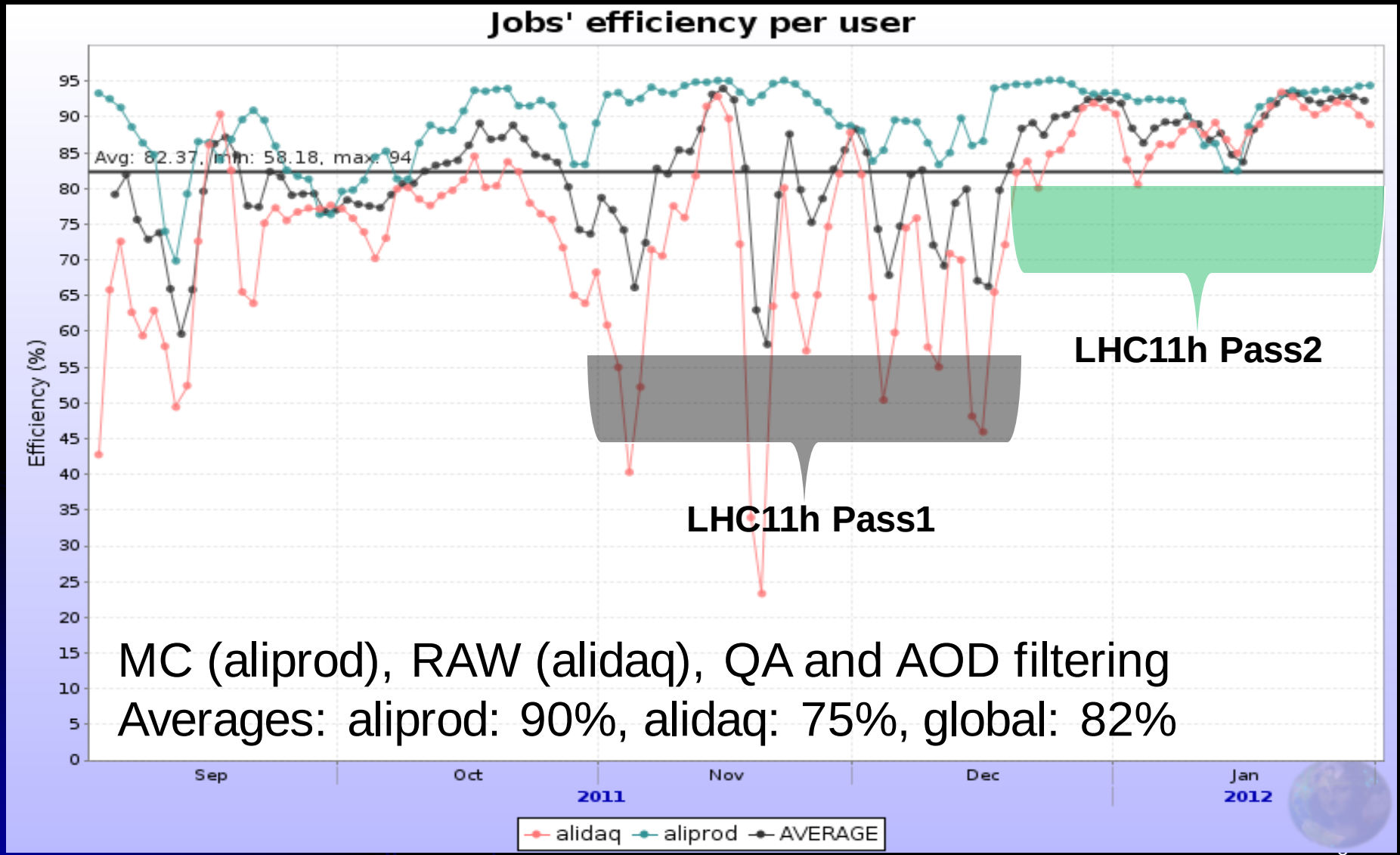
Grid Deployment Board

March 21, 2011

The 2011 PbPb data processing

- Second pass of LHC11h (PbPb period) completed
  - In time for “Hard Probes”
  - Still high memory issue (3.5-4GB), helped by the centres setting up ‘high memory’ temporary queues
- Third pass: possible before QM’2012 in August
- Presently running large MC productions
- ..and user analysis – more in this later

Production efficiencies



Analysis efficiencies

- Still need serious improvements
- Average ~30%, large variations
  - I/O intensive jobs are not ‘inefficient’ they simply use other aspects of the hardware…
  - … that said, the CPU/Wall is what is being measured and reported and thus we work to improve it
  - Efficiency strongly depends on the type of analysis

Analysis efficiencies

The LEGO Analysis Trains

- Polling together many user analysis tasks (wagons) in a single set of Grid jobs (the train)
- Managed through a web interface by a Physics Working Group conductor (ALICE has 8 PWGs)
- Provides a configuration and test platform (functionality, memory, efficiency) and a submission/monitoring interface
- Speed – few days to go through a complete period

MonALISA

Web interface
LPM

AliROOT

Analysis
Framework

AliEn

Grid jobs

Status

- Trains created for 6 PWGs
  - 4 PWGs already actively submitting trains
  - 2 in testing phase
- Gains
- Up to 30 wagons (individual tasks) for the most advanced
- Even for non-optimized trains the average efficiency is >50%

ALICE Analysis Trains Welcome **jgrosseo** - Help

PWG	Train name	I'm in	Last run	Description	Train operator(s)
CF	CF_PbPb	✔	09 Mar 12	Train for data PbPb running	jgrosseo, miweber
CF	CF_PbPb_MC	✔	09 Mar 12		jgrosseo, miweber
CF	CF_pp	✔	24 Feb 12	Train for AOD pp correlation analyses	esicking, jgrosseo
CF	CF_pp_MC	✔	18 Feb 12	Train for AOD MC pp correlation analyses	esicking, jgrosseo
GA	GA_PbPb		07 Mar 12		mcosenti
GA	GA_PbPb_MC				mcosenti
GA	GA_pp				mcosenti
GA	GA_pp_MC				mcosenti
HF	D2H_PbPb		10 Mar 12	D2H train for PbPb data analysis	jgrosseo, zconesa
HF	D2H_pp			D2H train for pp data analysis	jgrosseo, zconesa
HF	Electrons_PbPb			train for HFE PbPb	ssakai
HF	Electrons_pp				jgrosseo, sma
HF	Muons_PbPb				cheshkov
JE	Jets_PbPb		04 Mar 12	Jet analysis train for 2010 PbPb data	jgrosseo, kleinb, mverweij
JE	Jets_PbPb_2011		03 Mar 12	Jet Train for PbPb 2011	kleinb, mverweij
JE	Jets_PbPb_AOD			Train for Jet Analysis on AODs	kleinb, mverweij
JE	Jets_pp		09 Mar 12	Train for jets in pp	kleinb, vajzerm
JE	Jets_pp_MC		11 Mar 12	Jet train on pp simulated data	jgrosseo, kleinb, vajzerm
LF	LF_PbPb				delia, mnicassi
LF	LF_PbPb_MC				delia, mnicassi
PP	QATrain				mgheata

Future of trains

- We see that as the only viable method for running large scale (many hundreds of TB input data) analysis on a regular basis
  - ~300TB is what the users have to go through today to analyze the 2010 and 2011 Pb data samples + associated MC
- Works equally well with smaller samples – the time gain is significant over chaotic user analysis
  - Does not replace chaotic fully, but no need to run over everything anymore…

Storage

- ALICE has EOS – and we are ***extremely*** satisfied with the performance
  - It took ½ hours (4-emails) to configure and put in production – many thanks to IT/DSS
- New xrootd v.3.1.0 is being deployed at all SEs
- Following a T1-T2 workshop in January – sites call for a significant improvement of SE monitoring and control
  - xrootd development is ongoing
  - new MonALISA sensors for the servers

VO-boxes

- Re-write of proxy-renewal service (IT/ES – M.Litmaath)
  - Already running at several boxes
  - Clearly more stable, eliminates the room for error in handling the proxies
  - To be deployed gradually at all sites
- Reduction of number of AILCE services
  - SE already eliminated (last year)
  - PackMan (legacy) not needed with torrent
  - All of this is in AliEn v.2-20

AliEn v-2.20 upgrade

- New version ready
- New catalogue structure (lighter)
  - Presently @346.8 M PFNs, 2.5x LFNs
  - Growing at 10Mio new entries per week
- Extreme job brokering
  - The jobs are no longer pre-split and with determined input data set
  - Potentially one job could process all input data (of the set) at a given site
  - The data locality principle remains
- The site/central services upgrade on the fly – no downtime

SLC6 notes – from the web…

- Many applications will see a large increase in virtual memory usage (but not resident memory) when running on RHEL 6 vs RHEL 5. For example, see the following "not a bug" bug
- glibc 2.11 which includes a new "per thread arena memory allocator."… especially relevant for applications using many threads
- Note that this is all virtual memory, not committed/resident memory. An application should not use appreciably more resident memory than it did under RHEL 5.
- May have implications on sites batch system limits

Services summary

- **NTR** (as per daily operations meetings)
- Minor hiccups, very good support at all centres
- Disk is critical
  - Continuous cleanup of replicas (of older productions)
  - Priority to full storages and T1s
  - Eagerly anticipating the installation of 2012 pledges, already done at some sites

Summary

- The 2011 HI period data is processed and being analysed
- Ready for the 2012 data taking

- Many thanks to the T0/T1/T2 centres for the excellent performance and support

- ALICE is happy ☺