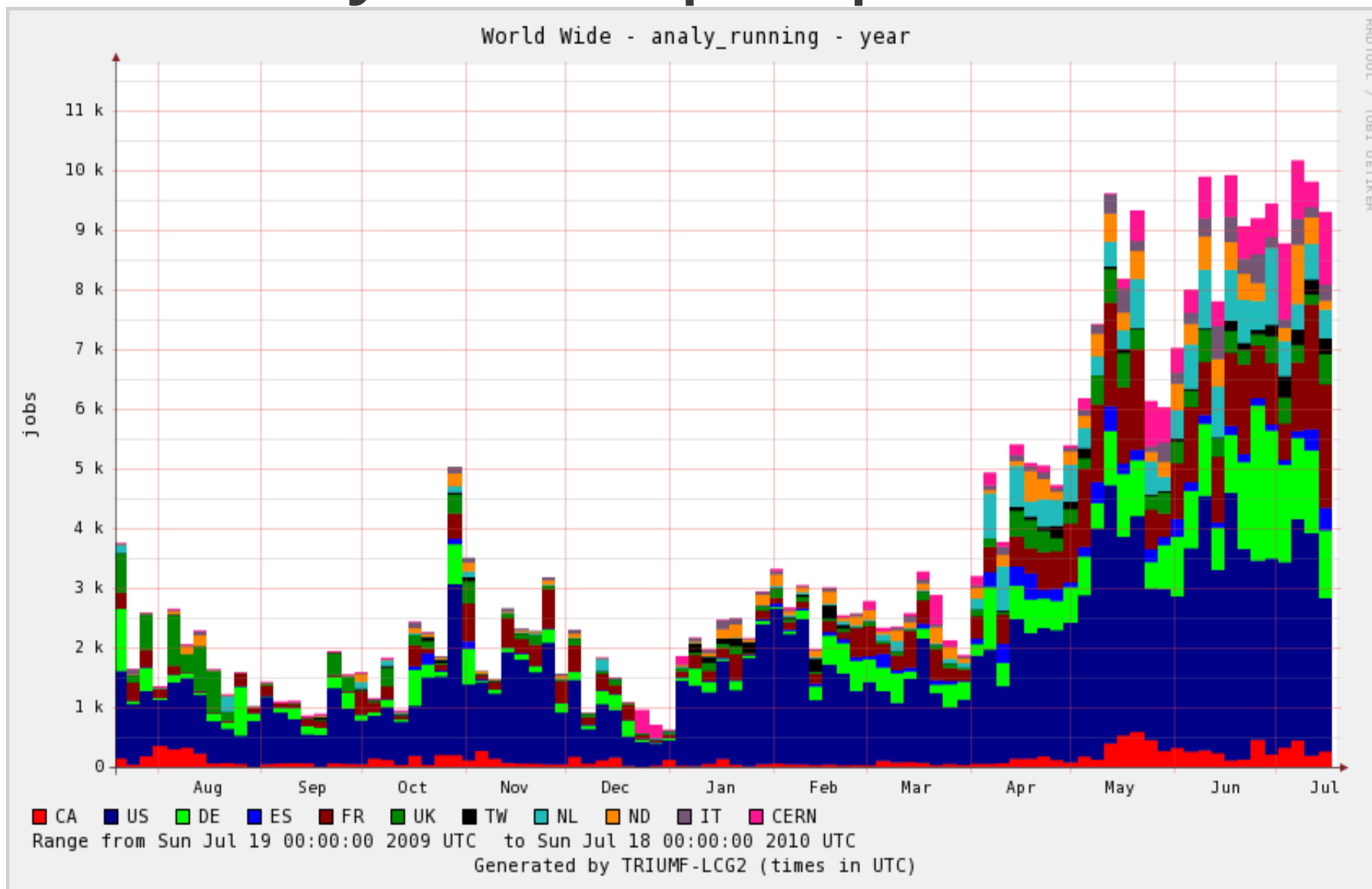


User Analysis Issues

- Rod Walker, 19th July 2010

- Ramp-up
- Data access methods
- DE specifics
- User problems/problem users

Analysis ramps up: Panda



Analysis Pattern

- User analysis pattern on Grid is changing
 - MC volume is small, off-grid analysis possible
 - data must be analyzed on the grid
 - non root-tuple formats already too big for local disk
 - BNL != Grid
 - long queue for users specifying BNL explicitly
 - all other clouds have short queues and good efficiency
 - including CERN
 - still much analysis runs at T1s
 - need to push most to T2s.

Data access methods

- Copy-to-Scratch
 - input size limited by WN disk ~14GB
 - io performance degrades as jobs share node
 - once staged, job independent of SRM
- File Stager
 - copy-to-scratch from athena, delete, copy next
 - no input size limit
 - io performance degrades as jobs share node
 - worse with simultaneous read/write?
 - not independent of SRM
 - fresh file in buffer cache – not for full host

Data access methods(2)

- Direct io, via dcap,rfio,file
 - no input size limit
 - independent of WN disk – scales with SE spindles
 - not independent of SRM
 - buffer cache not used
 - libdcap++ does good job too
- **This is DEFAULT for DE cloud**
 - including DPM sites: should look at rfio performance

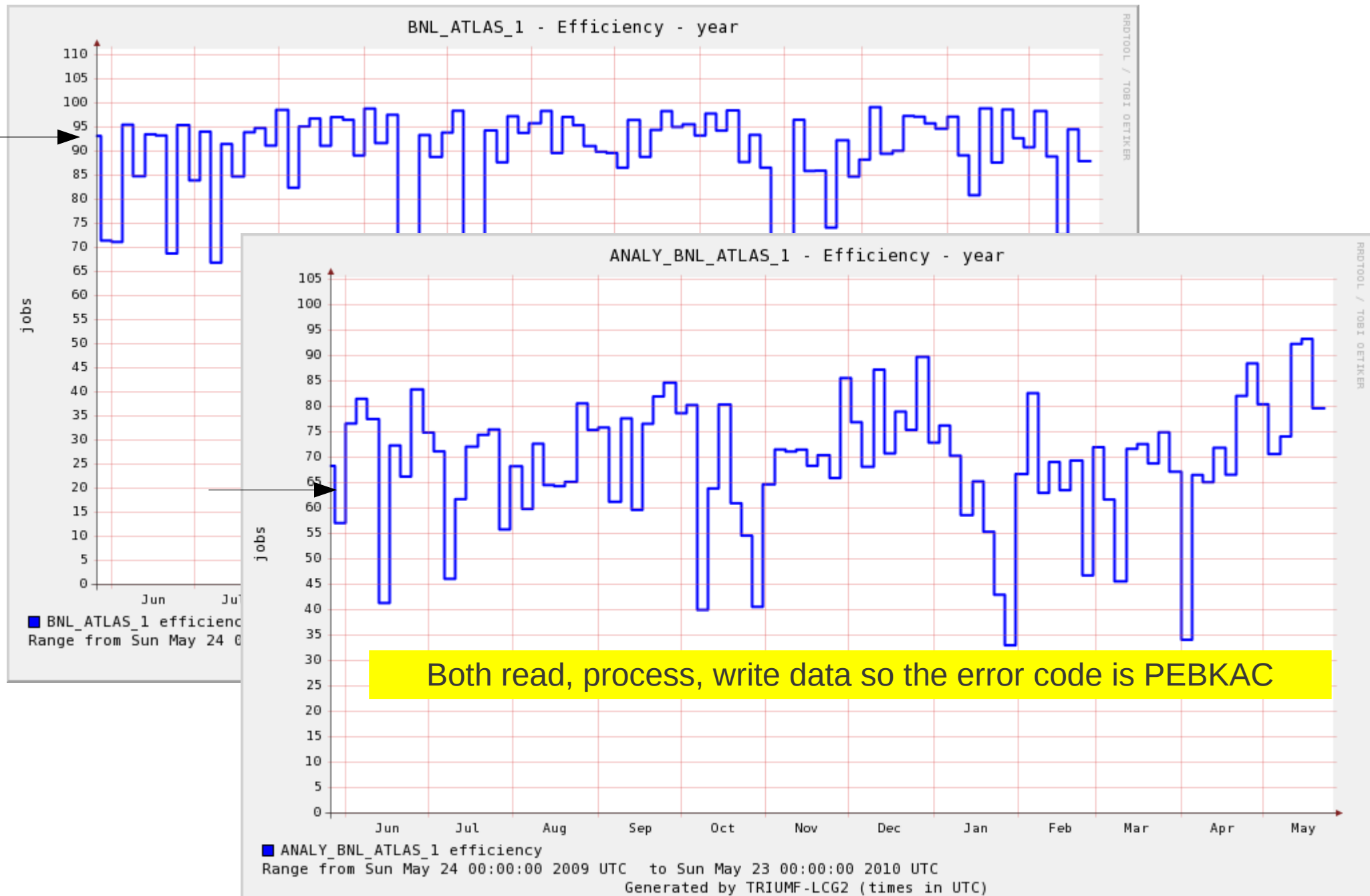
Access methods for users

- Ideally is transparent – it is not!
- Clients have various ways to choose splitting
 - pathena ignores maxinputsize for direct io sites
 - but allows user preferences
 - maybe ganga can be more helpful
- Longer jobs(more input) better for everyone
 - site efficiency, user turnaround & fairshare
- Direct io only works for root files
 - DBRelease, code tarball copied to scratch
 - ByteStream must be copied too(forgot this)

DE user bonus

- Several sites have extra resources for DE users AND configured BS accordingly
 - GridKa, DESY-HH, Freiburg get pilot streams which *only* pick-up DE user jobs
 - DE user = /atlas/de voms group
 - jobs submitted with this group will be picked up by *both* normal *and* DE pilots
 - maximize your share and throughput
 - jump 20k queue at FZK

User vs. prod efficiency



Problem Exists Between Keyboard And Chair (PEBKAC aka. ID10T)

- Grid/site/data problems clearly still a factor
 - reduced due to ProdSys spin-off
 - Adcos, DAST and P1 shifts
- User interfaces have hidden many Grid complexities
 - but cannot anticipate everything
 - tips for DE grid users ...

Top 10 user mistakes

- Using recon.AOD instead of merge.AOD
 - recon is only on single T1 and will be deleted
 - merge dataset has many copies on Grid
- Using ESD instead of DESD
 - DESD has multiple copies, ESD often only 1
 - please share use-cases requiring ESD
- Cannot find geometry version
 - Geometry only in DBRelease(not Frontier)
 - DBRelease in base release is old
 - LATEST(ganga keyword) has all previous Db tags
 - default now for both clients - SOLVED!

Top 10 (cont.)

- 'Testing' with 1000's of jobs
 - run over complete input dataset
 - every run jobs fails – test first, with 10 jobs
 - WOL defence (Works On Lxplus) not sufficient
- 7GB output limit per job(was 5GB)
 - Jobs will fail - WN disk space is limited
 - SCRATCHDISK is also not unlimited
- Very short jobs
 - grid overheads and initialization dominate wall time
 - sometimes limit from input size (copy-to-scratch)
 - remote-io coming soon to a T2 near you

Top 10 (cont.)

- SW version affinity
 - cannot keep your favourite version indefinitely
 - physical SW area space
 - support – bug response will be 'update'
- Prevent jobs splitting between sites
 - input dataset container, then ALL N datasets must be on the same site, i.e. forces T1
 - submit N jobs, then some will run at T2
 - problem of output aggregation solved in few weeks: output container dataset
 - recommend manual split and aggregate for now

Conditions Data Access

- 3 parts: GeomDB, Frontier, Flat payload files
 - Local sqlite; multiple failovers; **problem**
- Flat files distributed via DDM
- PoolFileCatalog has direct io TURLS to flat files
 - can be stale or corrupt
 - user jobs fail
- Need failover method for flat files
 - http turls in PFC with files on web server

Summary

- Access methods
 - direct io default in DE, and coming elsewhere
 - must check dpm rfio adequate
- DE group
 - use /atlas/de group to increase share
- User errors and issues
 - sweeping Grid complaints decreasing
 - ID-10T errors can be fixed
 - with education/wiki – not really
 - with client improvement – yes, e.g. LATEST
 - still storage and other Grid problems
 - e.g. PFC problem