ATLAS Computing Status and Plans

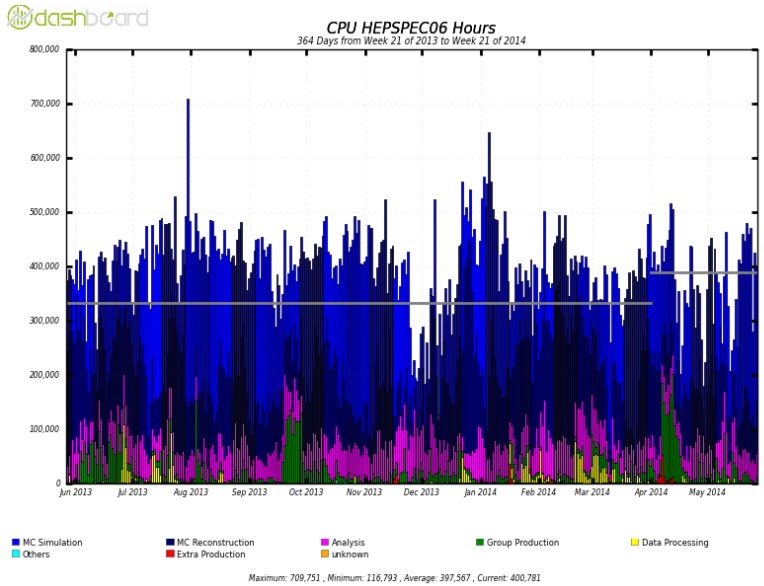
Richard P Mount
SLAC National Accelerator Laboratory

Topics

- Status – Resource usage in the last 12 months
- Focus on Distributed Data Management issues
- Plans – Preparation for Run 2

Status – Resource usage in the last 12 months

CPU Usage May 2013 to May 2014



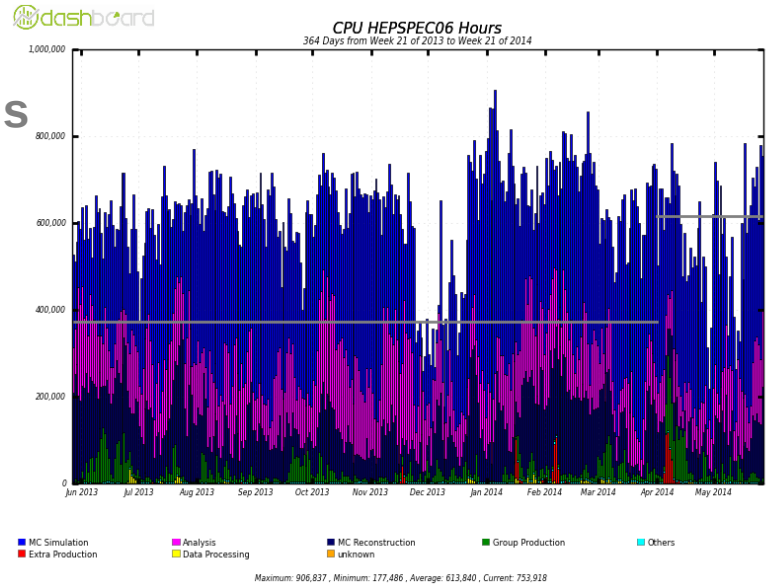
Tier 1s:

- Consistent above-pledge performance
- Saturation most of the time

MC Simulation
User Analysis
MC Reco
Group Prod
Group Analysis

Tier 2s:

- Consistent delivery of above-pledge and opportunistic resources
- Saturation most of the time

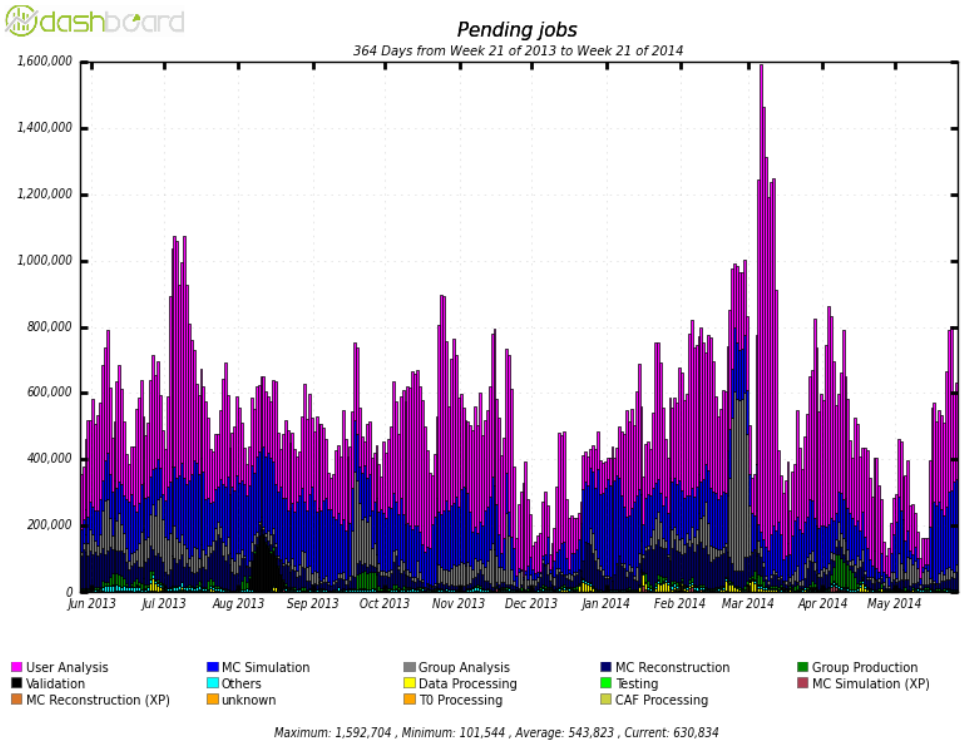


High Level Trigger Farm Exploitation

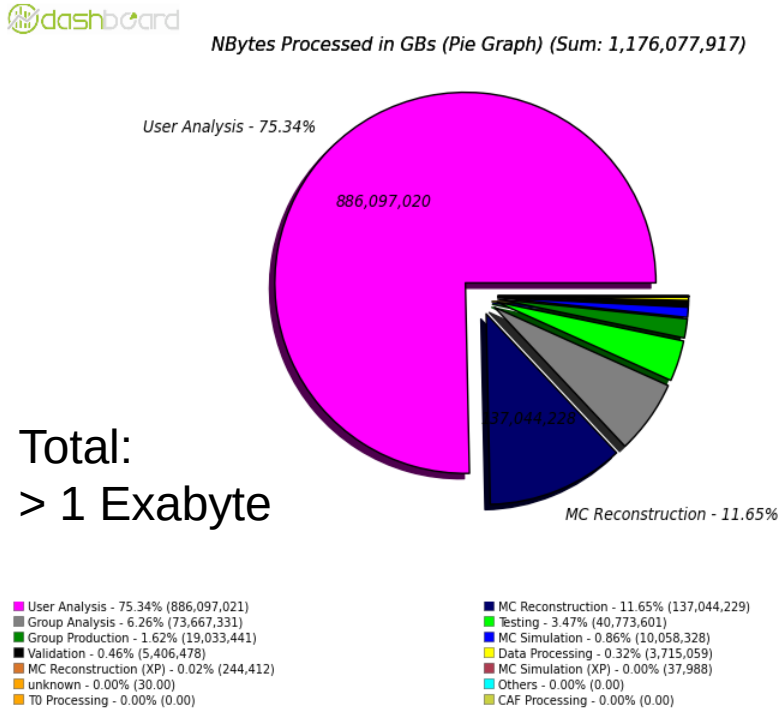
ATLAS HLT usage for grid jobs (bursts of over 20k jobs)

- The HLT has about 10% of the total ATLAS CPU capacity
- Its time-averaged availability for simulation is expected to be no more than 30% during Run 2

Pending Jobs and Volume of Data Processed



MC Simulation
User Analysis
MC Reco
Group Prod
Group Analysis

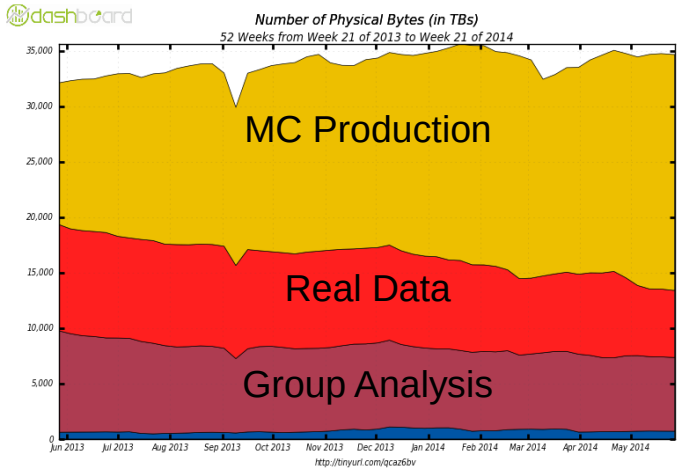


Must limit simulation to keep analysis turnaround acceptable

Analysis is the main driver of storage+network I/O capacity

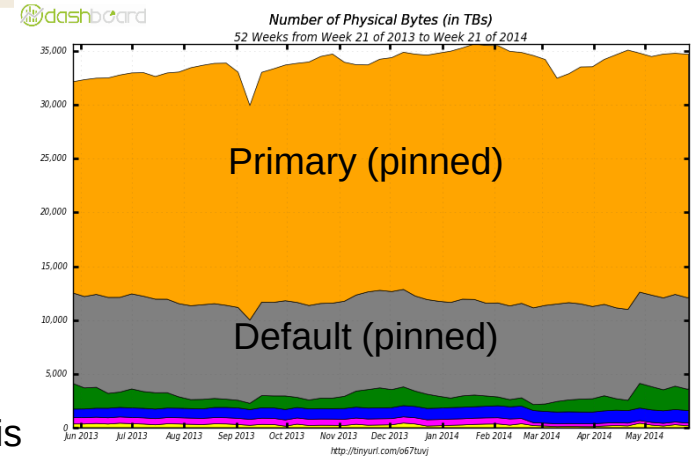
Total:
> 1 Exabyte

Disk Space



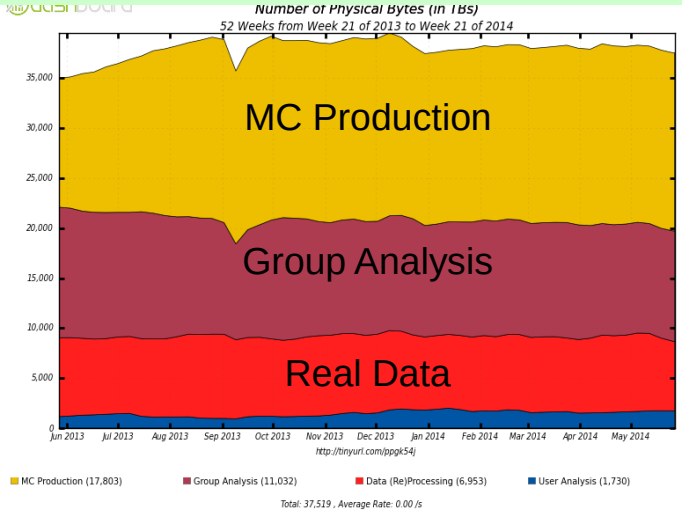
Tier 1

← User Analysis



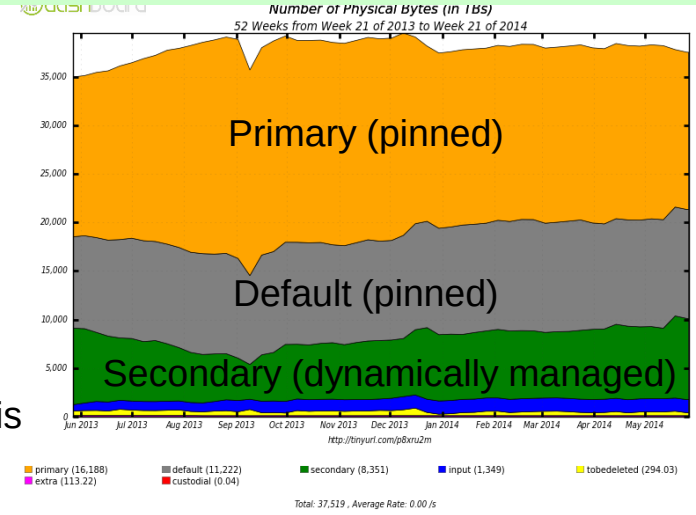
Secondary (Dynamically Managed)

T1 and T2 disks are full, requiring regular deletion of not-recently-accessed data
T1 dynamically managed space is unacceptably small (need to pin less data)

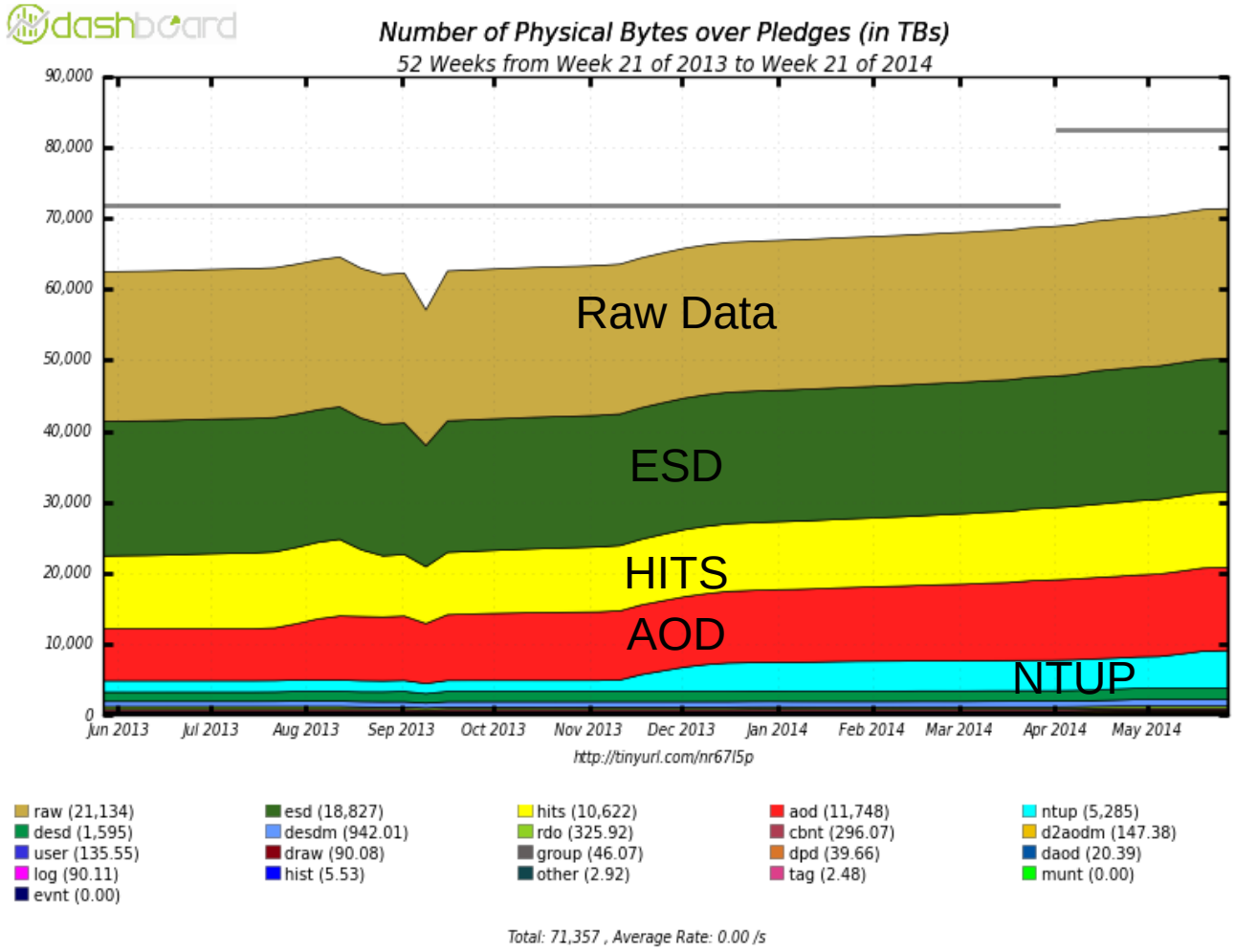


Tier 2

← User Analysis



Tape (Tier 1 + CERN)



Simulated Hits to be kept for ~1 year in future

ESD no longer written in most cases

Expect major growth of Group Data on tape.

Focus on Distributed Data Management issues

Computing Resources Scrutiny Group (C-RSG) – 1

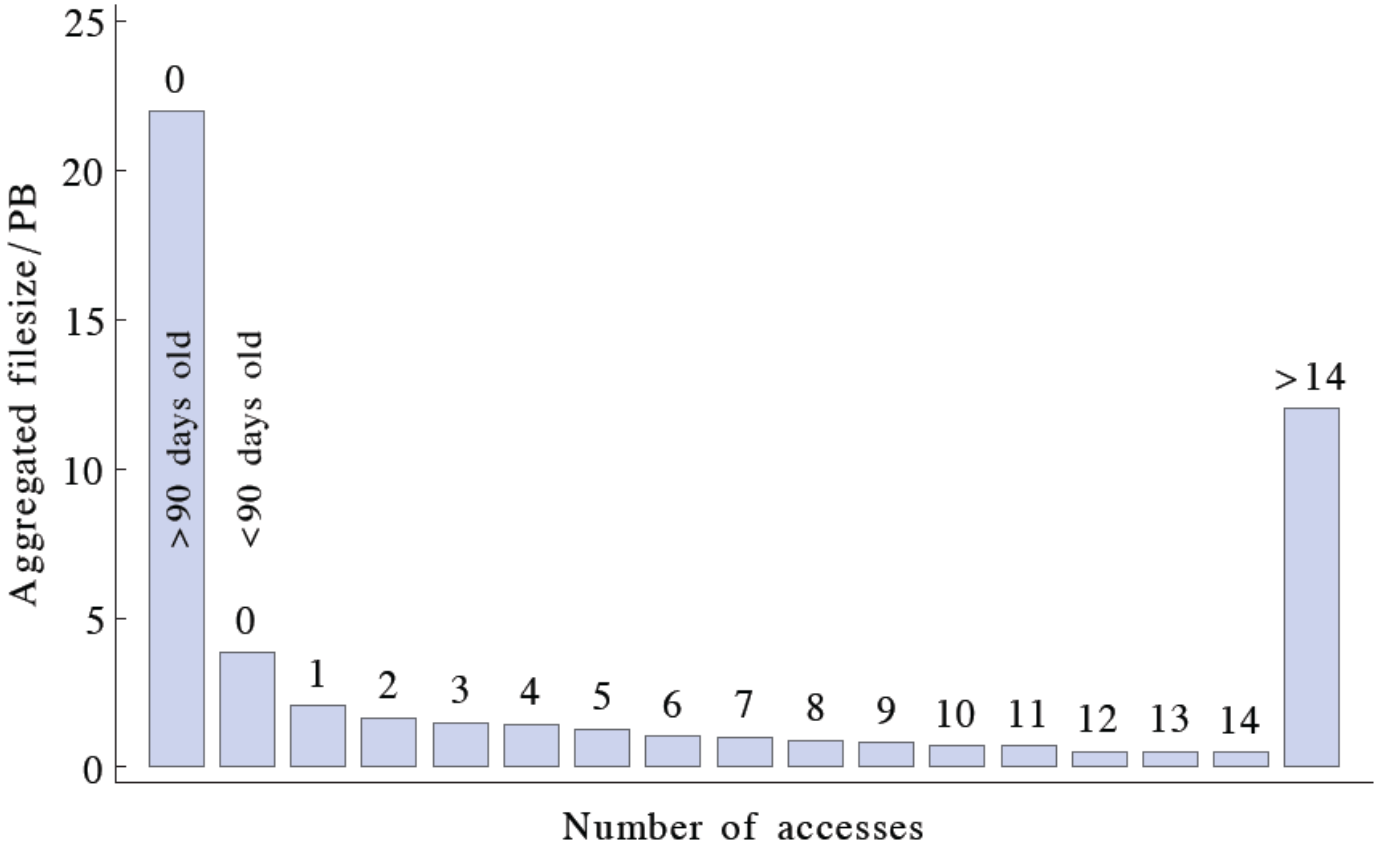


Figure 1 ATLAS DATADISK: volumes of data versus number of accesses in the 90 days ending 14 March 2014. Data created in the last 90 days but not accessed are in the second bin. The total volume of all DATADISK is 52 PB. Data supplied by ATLAS.

Unused Data

1. Recently created “production” data
   - Production tasks can take more than three months
   - Already seeing that some production output is being automatically moved to tape and deleted from disk before the production is finished.

2. Data for which there is no explicit lifetime or “move to tape” policy

3. Small-file data

The “unused data issue” is understood and is being addressed.

Towards a more dynamic disk/tape model

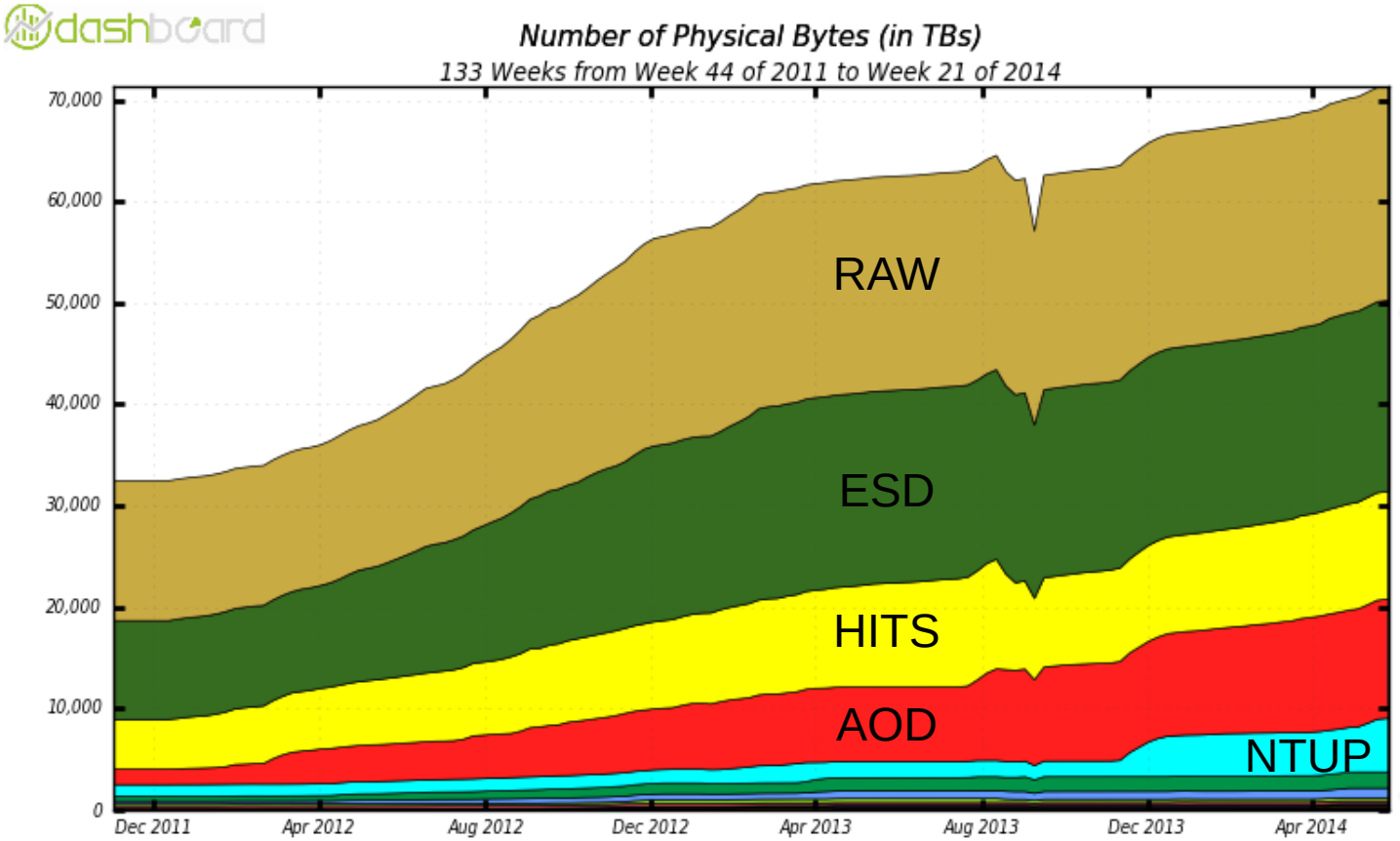
- All derived data will have a defined lifetime
- More systematic writing to tape except for transient data
- More automated deletion from disk when space is needed

Implications for tape systems

- More data will be written to tape
- Old tape space will be released (apart from raw data) on the timescale typical of tape re-packing at T1s

→ More data on tape and more writing and reading of tapes

ATLAS Tape Usage since 2011



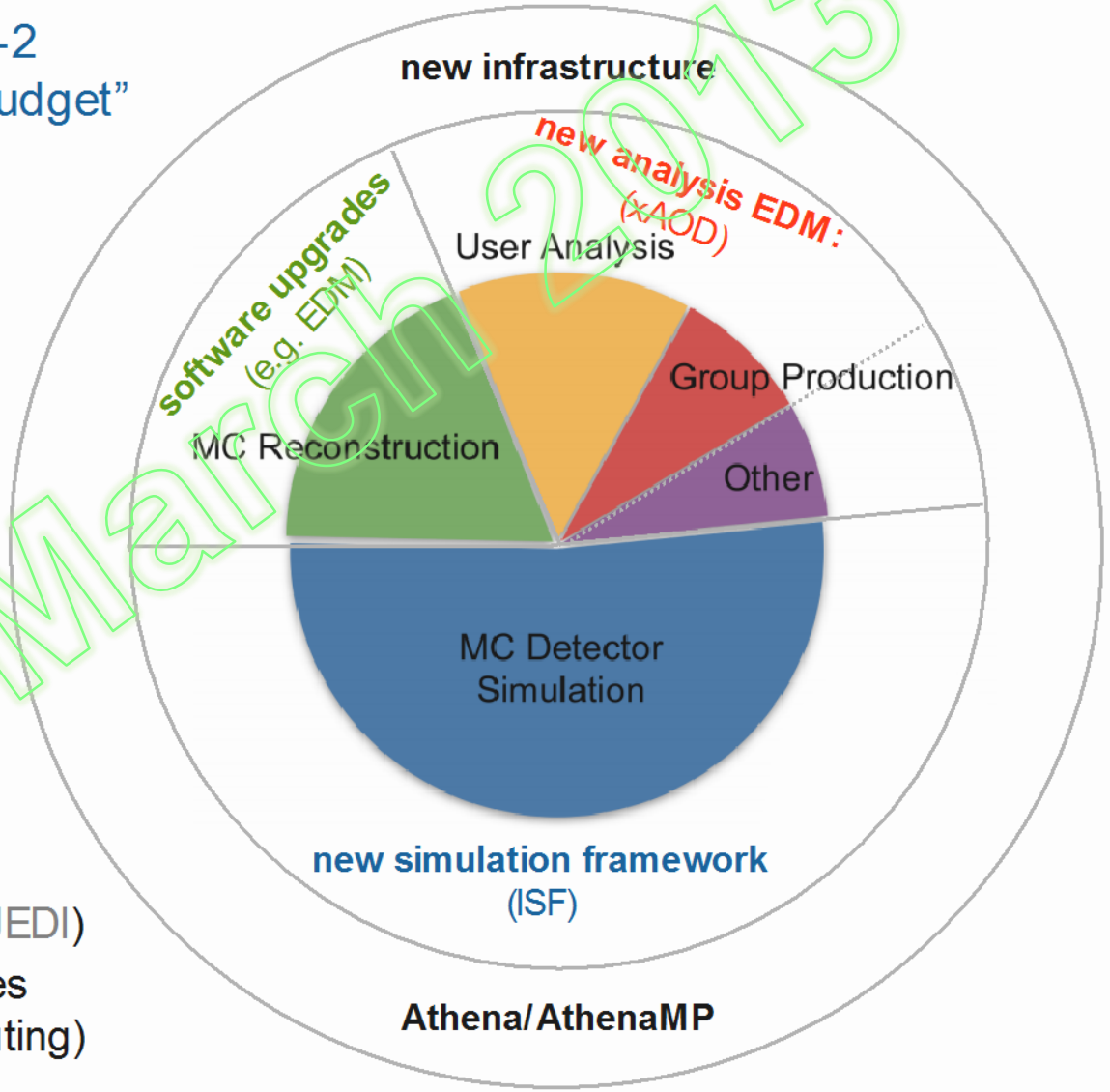
These data will have finite lifetime

→ Keep on tape for typically 2 to 3 years

Plans – Preparation for Run 2

Offline computing - preparation for Run-2

- Resource projections for Run-2 follows expectation of "flat budget"
  - update to 1kHz HLT rate
  - expected pile-up increase to 40
  - increased demands of MC statistics
- Need to optimise all across software and computing
  - CPU, disk size, workflow
- New GRID infrastructure
  - new data management system (Rucio) which scales beyond expected Run-2 data volumes
  - new workflow definition and job management system (Deft/JEDI)
  - exploring opportunistic resources (cloud/high performance computing)

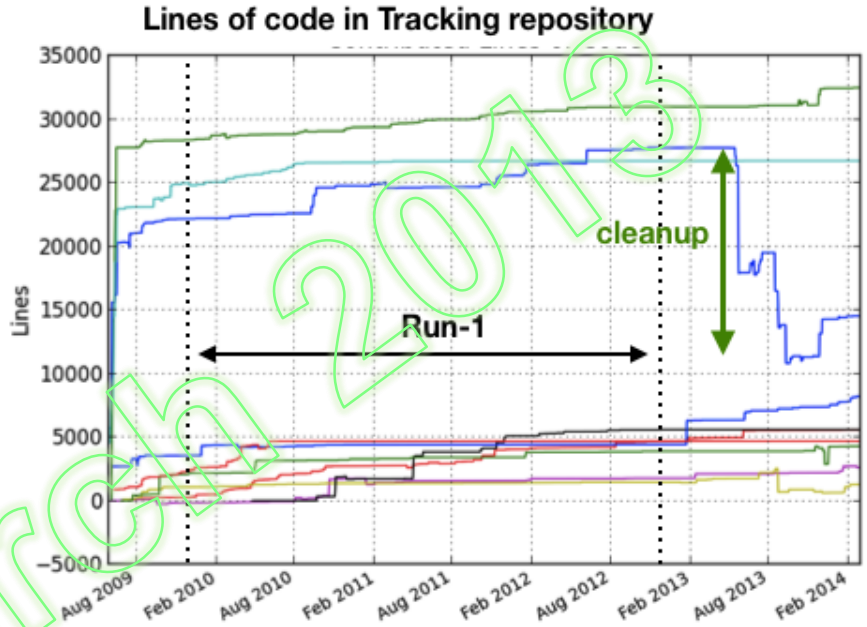


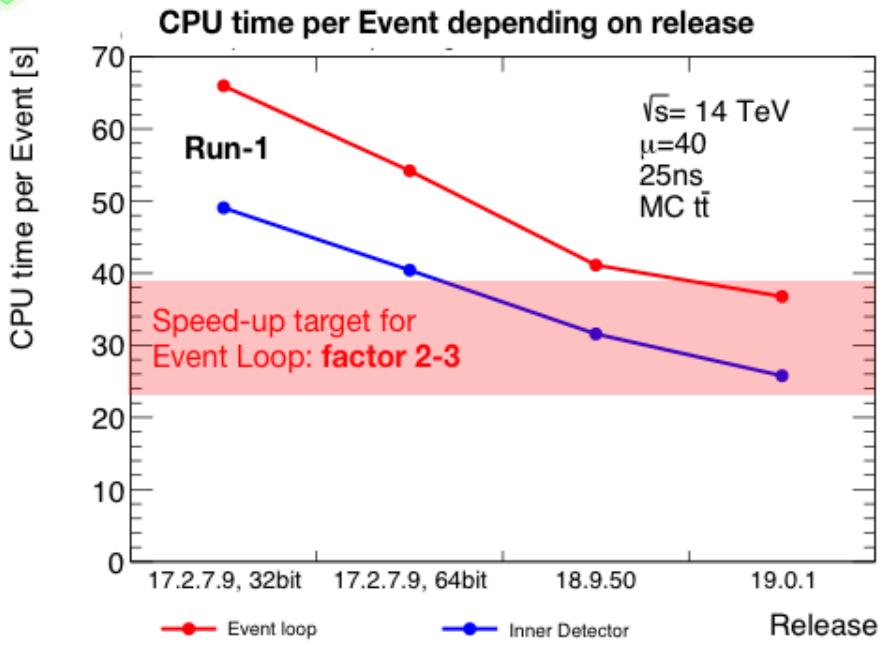
to be exercised in data challenge 2014 (**DC14**)

MC/group production and user analyses averaged over all ATLAS grid sites (T0/T1/T2/T3)

Optimising CPU needs

- Large-scale software cleanup and optimisation program on the way
  - `flat EDM` structure to remove overhead
  - replacement of algebra/geometry library, many alternatives to CLHEP tested, finally Eigen library chosen
- Main single CPU consumer reduced by **factor 2** compared to Run-1 release
  - accumulating changes from 32bit->64bit, new magnetic field service with enhanced caching, Eigen integration
- Replacement standard math library
  - candidates are (VDT, Intel)
- First release with new EDM (19.0.0) built late january
  > 1000 packages reworked





Reconstruction Status Today

Release 19.0.2.X is in final validation for DC-14 / 8 TeV

- on track to reach ~3 fold CPU speedup (target was 2-3) for Run-2, but optimisation still in progress
- expect definite CPU numbers soon, to be used for detailed Run-2 resource planning

DC-14 / 8 TeV will be first large scale production of xAOD according to new Analysis Model

- successful completion of major rewrite of reconstruction software to produce new format
- much work remains, especially to optimise xAOD size (right now disk space is likely to be the main resource limiting the physics precision for Run-2)

Summary

Resource usage in the last 12 months:

- Consistent usage at (disk, tape) or beyond (cpu) the pledge level

Distributed Data Management issues:

- Disk space constraints (almost no dynamic buffer at T1s)
- Not-recently-accessed data – we are recovering disk space but will use more tape I/O and space.

Preparation for Run 2

- Major improvements to Reconstruction, Simulation, Analysis and Distributed Computing