R&D projects Brainstorming

ATLAS Italia Computing

4 March 2021

Zach Marshall (LBNL) and Ale Di Girolamo (CERN)

Roadmap 2 Run 4 (R2R4)

- The ATLAS Computing CDR is now public (CERN-LHCC-2020-015)
- Projectizing of activities started as from LHCC recommendations
  - Iterative process to find a proper balance

- Strategy to the TDR:
  - Regular meetings with area coordinators to discuss the details of the R2R4 plan
  - Checkpoints internal to ATLAS to understand global status and interfaces between areas
    - each 9 months
    - next (first one) June 2021

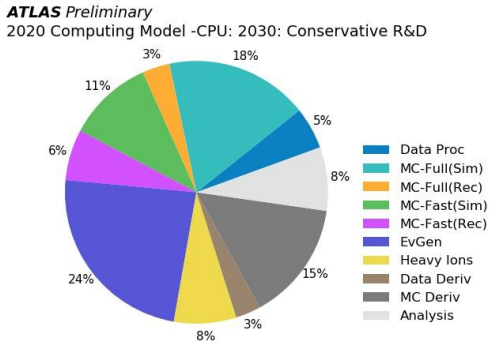
ID	Year	Milestone
P1.1	2020	CDR released: identifies HL-LHC R&D needs, and the projects attempting to address them.
P1.2	2020	Run 3 deployment: includes all Baseline updates and releases and potentially some Conservative R&D.
P1.3	2021	Run 3 starts: R&D focus shifts to HL-LHC R&D.
P2.1	2021	Run 4 R&D plans: all R&D projects targeting Run 4 provide a program of work to 2024, including risks and effort estimates.
P2.2	2022	Run 4 R&D demonstrators: all R&D projects targeting Run 4 provide proof-of-concept demonstrators
P2.3	2023	TDR released: prioritizes Run 4 Conservative R&D projects. Endorses a limited amount of Aggressive R&D projects for Phase 3
P3.1	2024	Run 4 projects approval: go/no-go decision for R&D projects targeting Run 4. Every project must provide functionally complete prototypes.
P3.2	2025	Run 4 development planning: each Run 4 project provides a WBS inclusive of effort and risk estimates. It also provides training and documentation tailored for the target developers community.
P4.1	2025	First Run 4 deployment: includes production-quality implementations of all new developments to be included in Run 4. Not all physics code migrated yet.
P4.2	2026	Second Run 4 deployment: includes production version of all Run 4 code and integration. Validation starts.
P4.3	2026	Run 4 dress rehearsal: test software and ADC readiness for data taking.
P4.4	2027	Run 4 starts: R&D focus shifts to Run 5.

Table 12: High-level milestones for HL-LHC Computing and Software R&D program

Roadmap 2 Run 4 (R2R4)

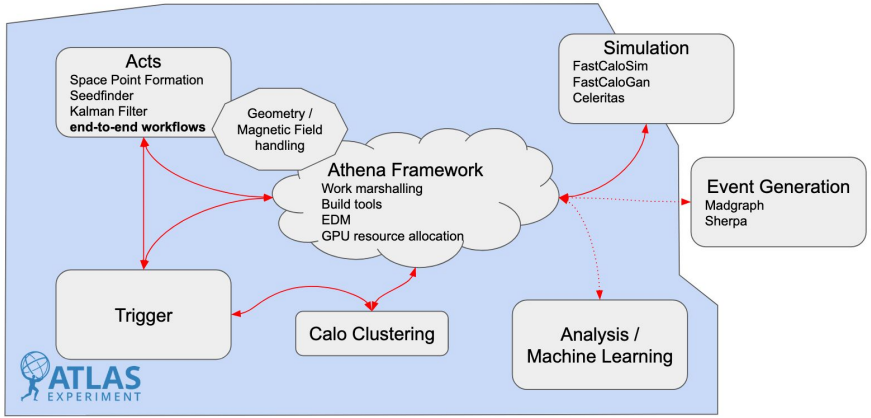
- Projectizing of activities progressing
  - Selected specific areas to start with, defined a structure to define R&D projects, with their impact and milestones.
  - First areas (Core SW and Simulation) already reporting plans.
  - Iterative process to find a proper balance.
- Steps towards the LHCC Computing second phase review
  - Regular meetings with WLCG PL, WLCG SW Liaisons and the others Experiments to refine the details.
  - Internal regular meetings with area coordinators to discuss the details of the R2R4 plans.
- Reporting in the next slides status and progress on a few specific challenging R2R4 areas
  - Heterogeneous Computing and Accelerator Forum
  - Storage Caches and Storageless sites

R2R4 - HCAF



- Estimated CPU needs 2030: not a single activity
  - But Event Generation and detector simulation will require most of the resources
    - Very important that we collaborate with the developers of these
  - Reconstruction and analysis: smaller piece of the pie, but under our control
    - In reconstruction, tracking and calorimetry are the first two targets for accelerators, in analysis it is machine learning
    - High-Level Trigger (HLT) reconstruction will profit especially from improvements in tracking

- **H**eterogeneous **C**omputing and **A**ccelerator **F**orum



Mandate for the Heterogeneous Computing and Accelerators Forum

(Updated on 14.1.2021)

Mandate:

The future of computing hardware is uncertain, but one global trend is towards heterogeneous resources and more specifically towards "accelerators": specialized (non-CPU) hardware that enhances performance for certain computations. One of the most obvious examples is the Graphics Processing Unit (GPU), which is adept at highly parallel, low-accuracy computations. Other popular examples include FPGAs and TPUs.

Within ATLAS, discussion and overall planning of work on heterogeneous resources should be within the Heterogeneous Computing and Accelerators Forum (HCAF) which includes efforts from both offline software and TDAQ. The conveners of the forum should maintain a list of high-level milestones towards the adoption of the technologies targeted by development within ATLAS.

The forum should meet at least once a month.

Reporting and Liaisons:

The HCAF conveners report to the ATLAS Computing Coordinator and the TDAQ Project, TDAQ Upgrade Project, and Upgrade Project Leaders. They may appoint liaisons or contacts as needed. They should ensure ATLAS is represented in collaborative forums focused on accelerators, like the HSF accelerators forum.

Term of Office:

The HCAF conveners are appointed by the ATLAS Computing Coordinator and TDAQ Upgrade Project Leader with a renewable one year term normally starting October 1st. At least two conveners are appointed. Between them, responsibilities are split; however, knowledge should be shared such that they can represent each other in case one is unavailable.

R2R4 - HCAF: Short and Long-term Plans

- Define portable parallelization strategies
  - “Parallelize once, run (almost) anywhere” (avoid lock-in where possible)
- Integrate and manage accelerators in ATLAS software framework
  - Memory allocation, resource scheduling
- Optimize accelerator utilization (helpful on and off accelerators)
  - Define accelerator-friendly data structures and algorithms
  - Process multiple events as a batch
  - Identify opportunities to use mixed precision

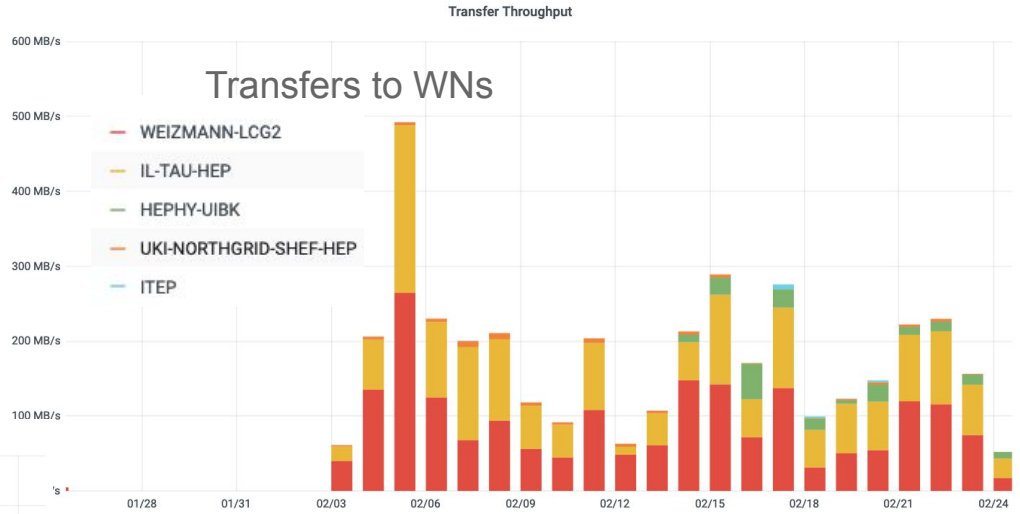
Cutting-edge R&D anchored by real-world examples

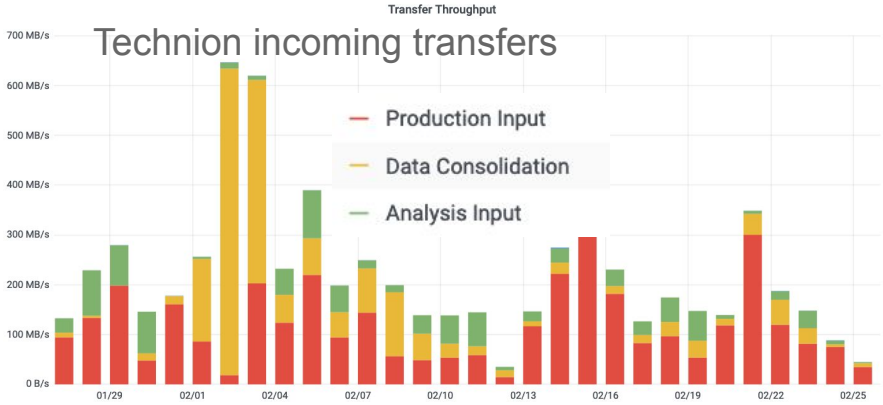
- Detector simulation, generators -- in collaboration with HSF, R&D projects
- End-to-end on GPU tracking -- in collaboration with ACTS, R&D projects
- ML algorithms-- in collaboration with ATLAS ML forum, R&D projects

R2R4 - Storageless Sites

- Since 2021 improved monitoring: now able to properly monitor data movement between grid storages and diskless sites
  - Critical to understand network usage

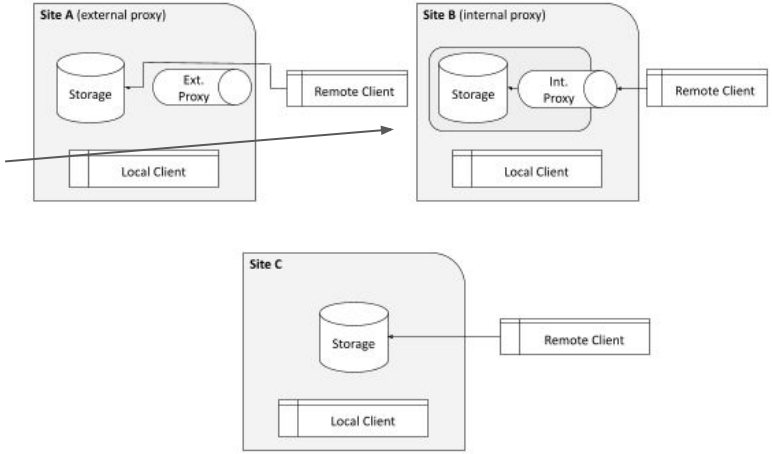
- Consolidation of storages, Israel example (3 sites, 1 storage)
  - WAN saturated by "Data Consolidation" + remote accesses of the other 2 sites
  - FTS transfers all failing → WN starvation
  - Manual tuning of max concurrent transfers





R2R4 - Caches

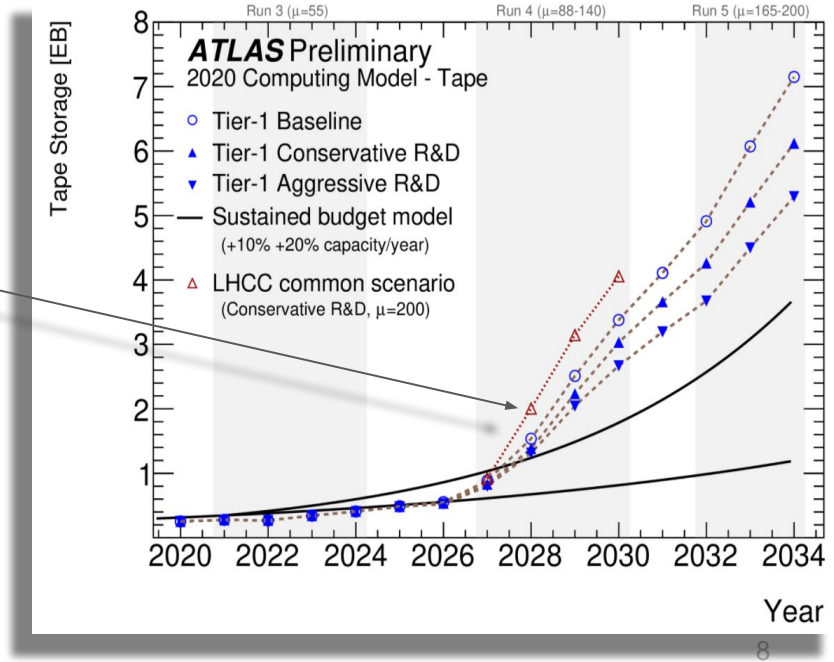
- Several sites deployed caches
  - Cache for some small CPU only sites
    - A few of them, e.g. in UK
  - At a few big T2s, in front of existing storage: R&D for analysis, creating mini-lakes
    - Positive experience, cache hit rate up to 80%, without visible performance degradation
    - Deployed with SLATE, a service which enables deployment to sites from service expert group



- Deployment and support model: possible evolution on same idea for other services
  - E.g. Squids; why not batch (kubernetes) in the future.
  - Concerns:
    - trust/security (despite all security concerns have been sorted out)
    - possible shift of responsibility from sites to central operations team

R2R4 - Data Carousel and tape challenges

- Data Carousel activities progressing
  - Details of 2015-2018 reported in the last LHCC
  - Now ongoing tests of new workflows (e.g. Derivations)
  - Second replica of AOD on tape to improve performance (reduce latency and tails)
- R2R4: huge challenges for T1s Tapes
  - In 2028, resource estimates of tape at T1s :
    - 2 EB: T1s Conservative R&D vs
    - 1.3 EB: 20% sustained bdgt model
  - And big throughput if we want to increase the tape usage to optimize disk storage
    - [here we can quantify better, slide in bckup]
  - We have started working to tackle these challenges →
    - Need coherent efforts from Tier1s and all experiments



R2R4 - Analysis

- DAOD_PHYS(LITE) is the baseline format for Run 3 (4) analysis
  - Perhaps this is not imaginative enough; do we need a ROOT format in Run 4? A file format at all (datalakes? databases?)?
- There remain a lot of interesting questions about how we do analysis in the future (late Run 3 / Run 4)
  - DAOD_PHYSLITE relies on *harmonization*, while physicists like *complex* analysis and interesting developments – are these in competition?
  - What about interactive analysis, notebooks, large (quasi-interactive?) resources (analysis facilities?)
  - What about systematics? Do we apply these before hand (better for notebooks?), on the fly (with a simplified tool?), some other way?

R2R4 - Analysis (Facilities)

- How will Analysis work in Run4?
  - As mentioned in the previous slide, a lot still to be shaped
- AF: buzzword or reality?
  - A bit of both
  - Scalable automatic deployment of services depending on the workload needs?
    - E.g. exploiting kubernets
  - Notebooks plus batch?
  - High throughput disks?
  - Where are our real limitations?
  - What do we need to speed up?
  - Usability?
- There is room for ideas and R&D projects here
  - Not just buying HW
  - Understanding our (present and future) workflows
  - Define key metrics for (analysis) success

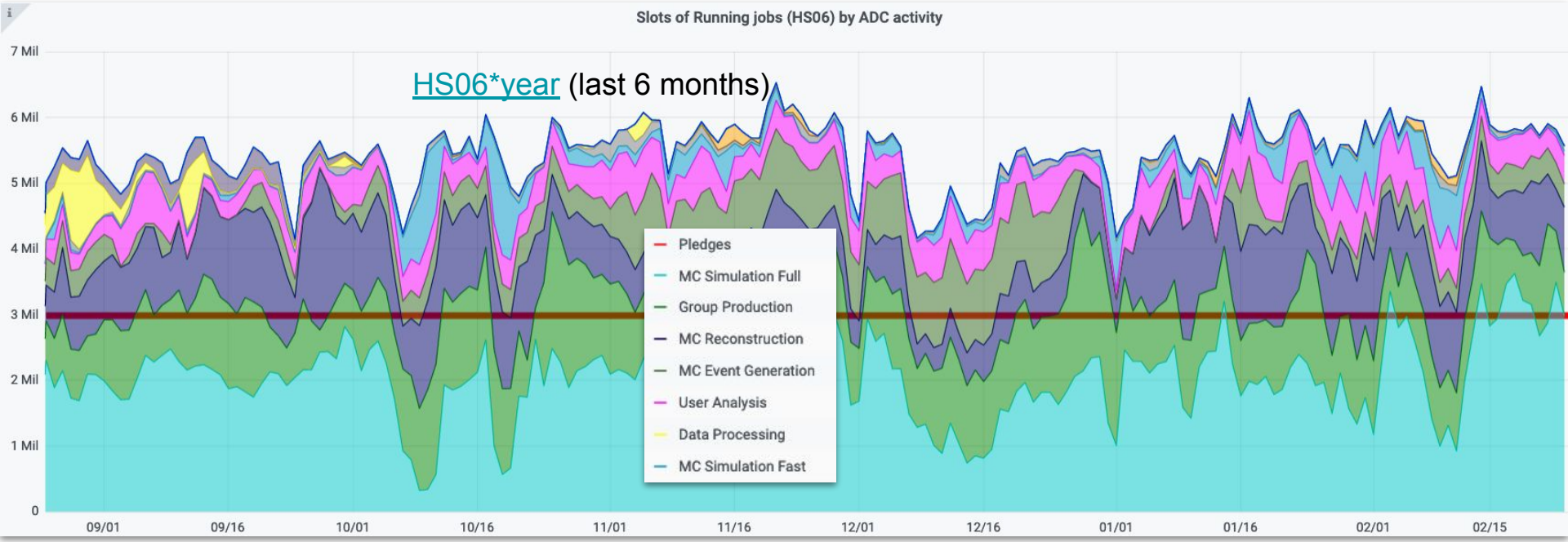
Summary: the present is shaping our future

- We need to engage today:
  - We are big; we need several years to develop, integrate and validate new ideas and technologies
  - Plenty of challenging R&D projects in front of us:
    - We need Institutes and FAs participating (proposing, leading!)
    - R&D projects will be on top of getting things working.

Backup

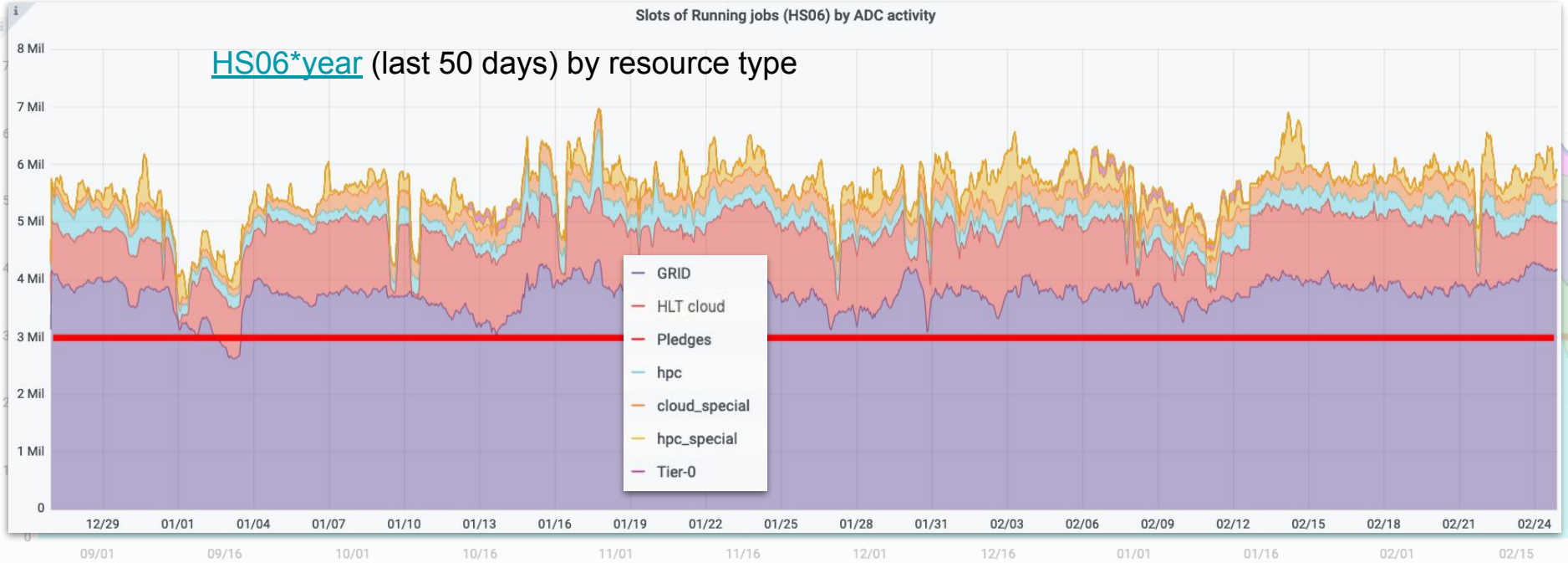
Resource usage - compute - overview by activities

- Exploiting well all the available resources
  - No major issues: 24x7 (best effort) operations work to anticipate & minimize the problems



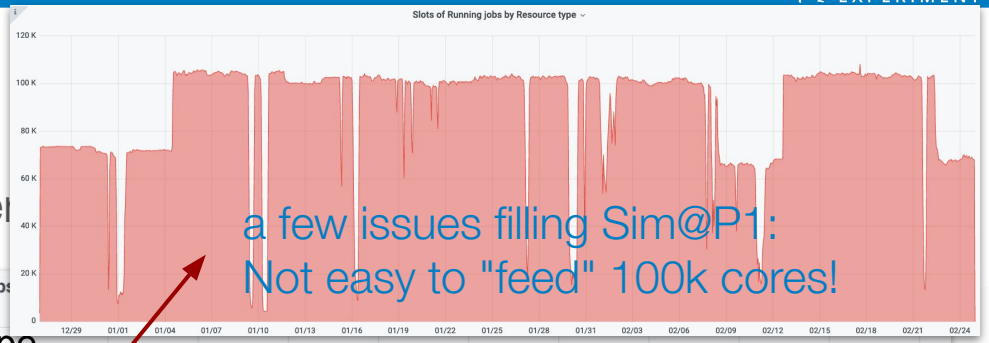
Resource usage - compute - overview by resource type

- Grid, Sim@P1 (P1 HLT farm), BOINC (Cloud_special), HPC (simple) and HPC Special
  - Exploiting a variagate set of different types of resources

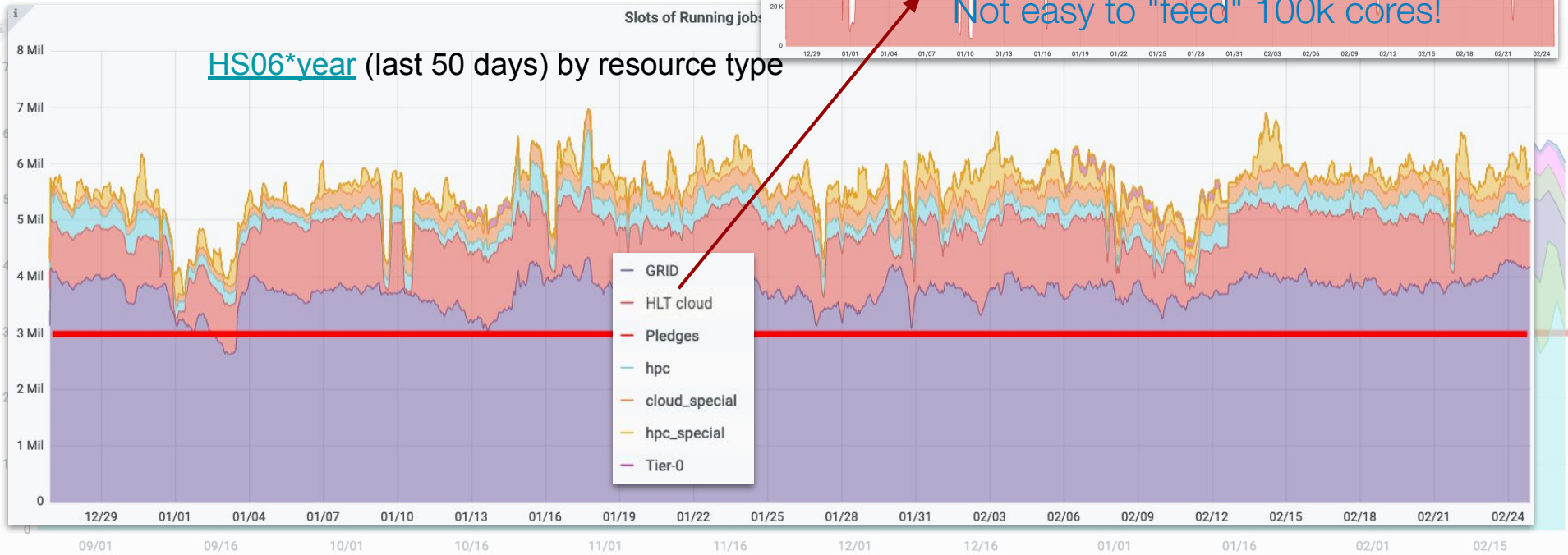


Resource usage - compute - zoom on Sim@P1 HLT farm

- Grid, Sim@P1 (P1 HLT farm), BOINC
  Special
  - Exploiting a variagate set of differ

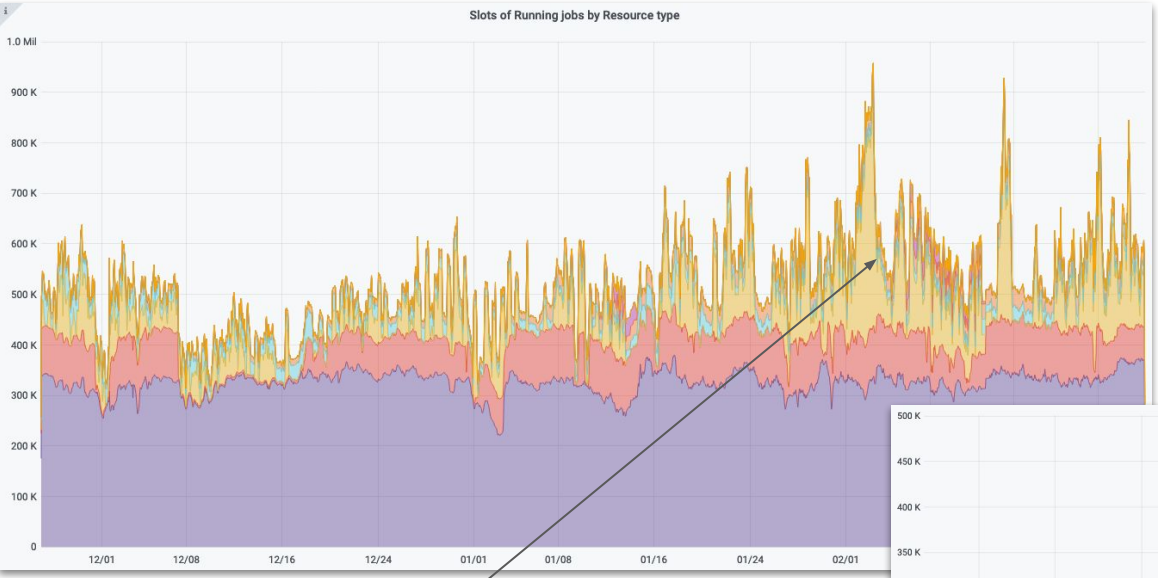


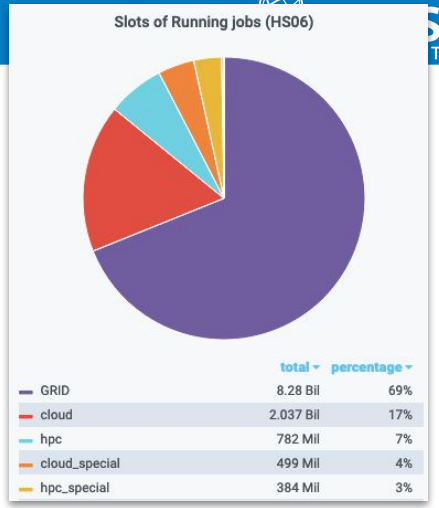
a few issues filling Sim@P1:
Not easy to "feed" 100k cores!



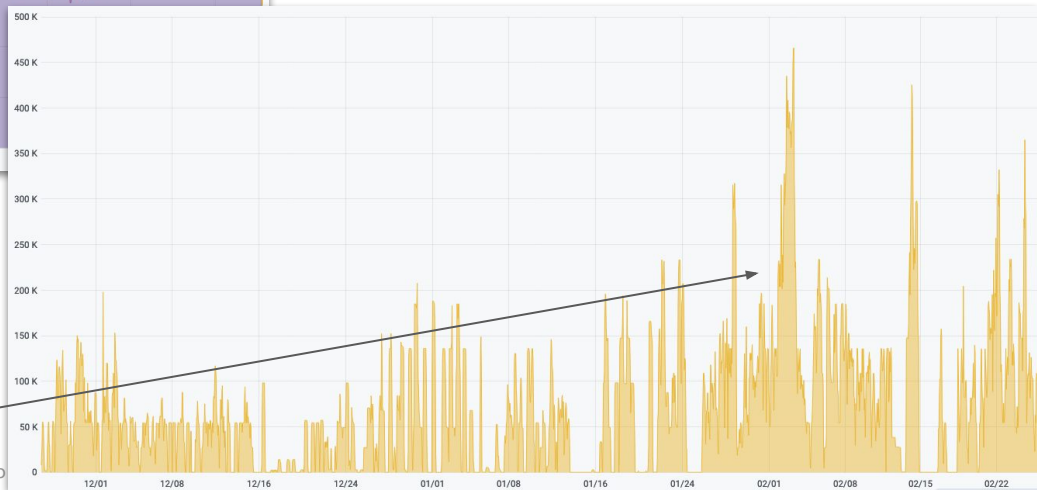
Resource usage - compute - HPC view

- HPC integrated and successfully running real production jobs



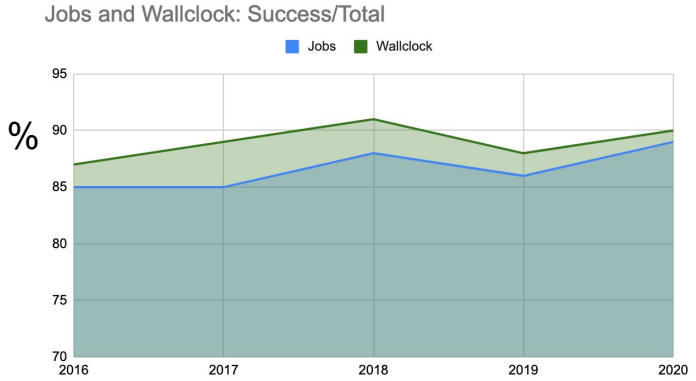


Super challenges!
Nothing is for free: evolution of our systems to double the capacity to absorb these peaks

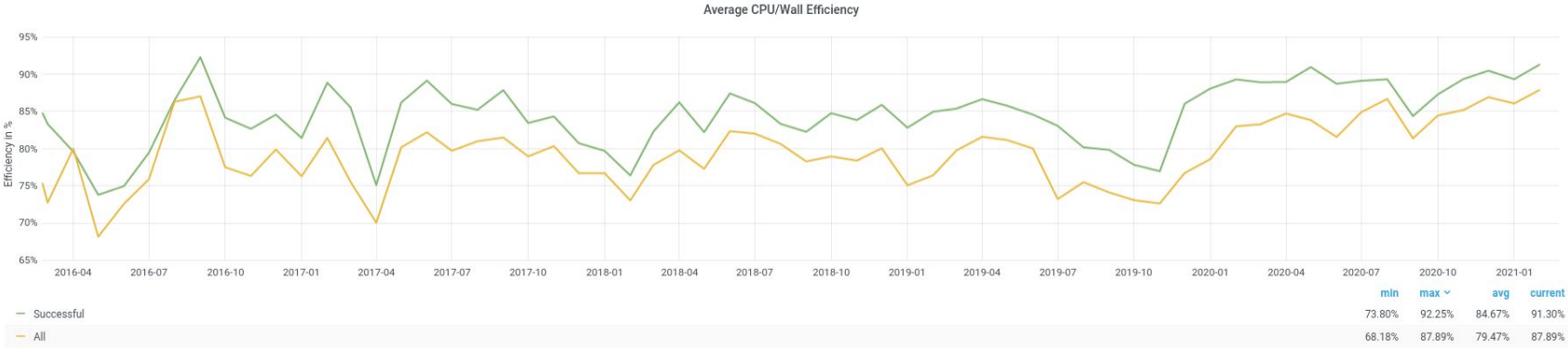


Resource usage - compute - Errors and Efficiency

- Successful Wall clock / Total Wall clock:
  - Constantly looking into the details to optimize and minimize wasted resources
  - Over the past 5 years ~ doubled the Wall clock, still effectively managing to keep the failed wall clock under control!
  - ~ 10-12%: approx 50/50 split between infrastructure and application issues

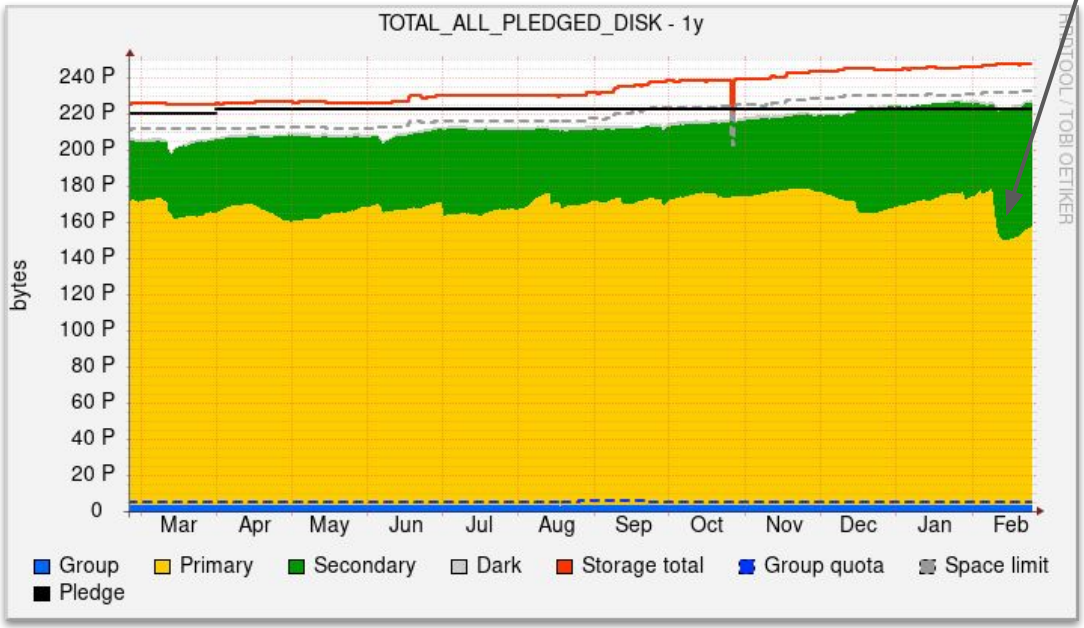


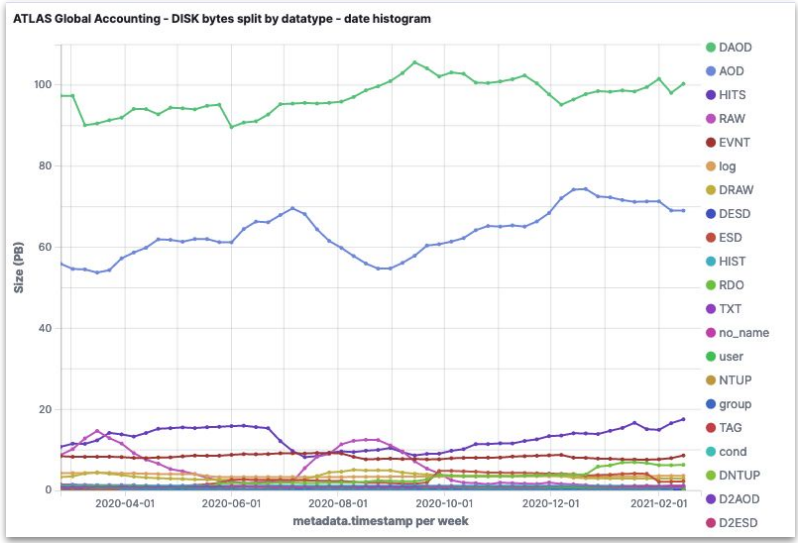
- CPU / Wall clock:
  - Healthy ~ 85-90%. As expected, mostly due to initialization and I/O



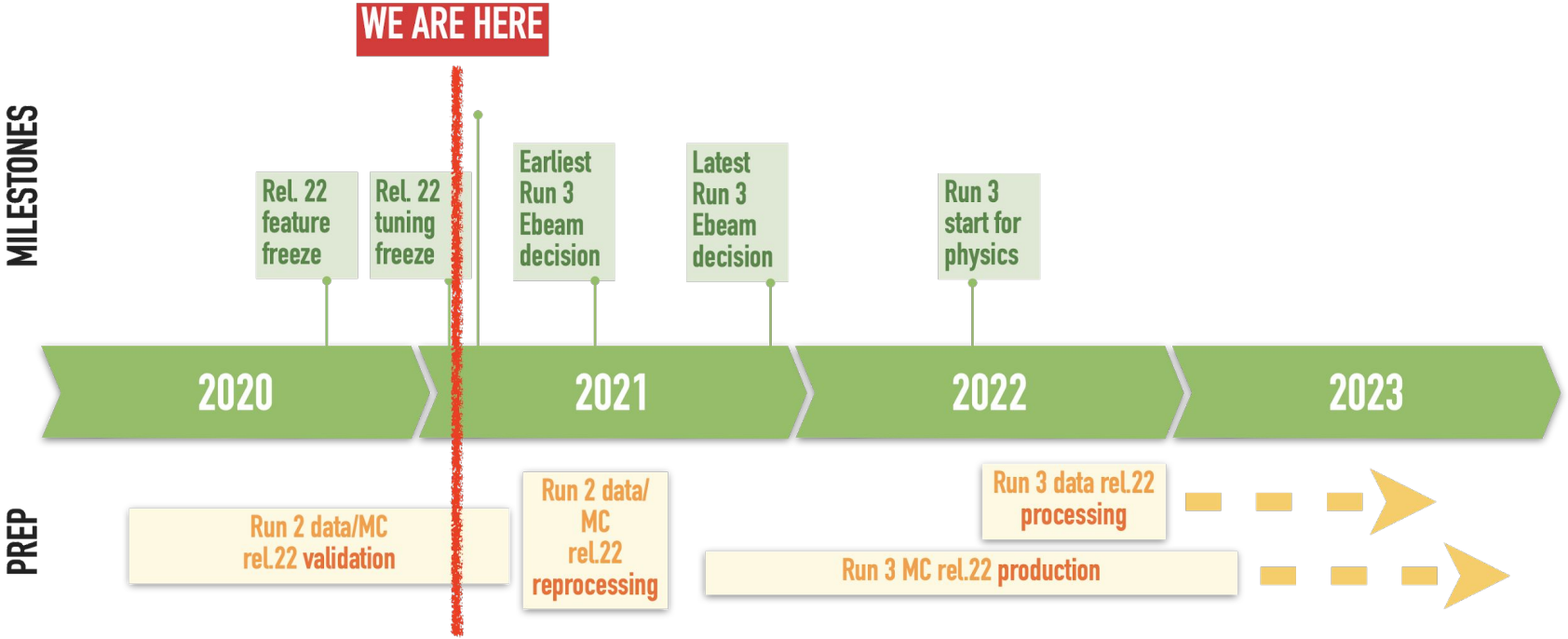
Resource usage - disk - last year

- Fully utilizing the pledged disk
- Secondary/total will improve:
  - AOD will no longer be permanently kept on disk, popular data will stay and less-popular will be recalled through data carousel. The gain is order of 20PB of primary disk space





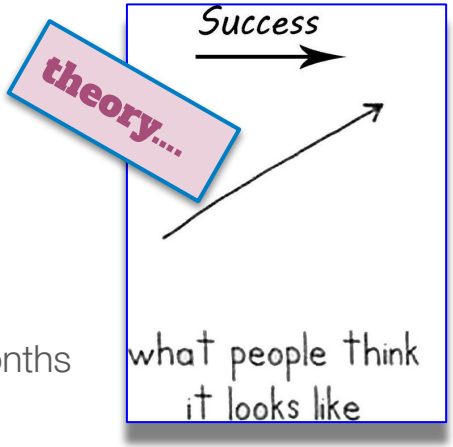
Release 22 – AthenaMT status

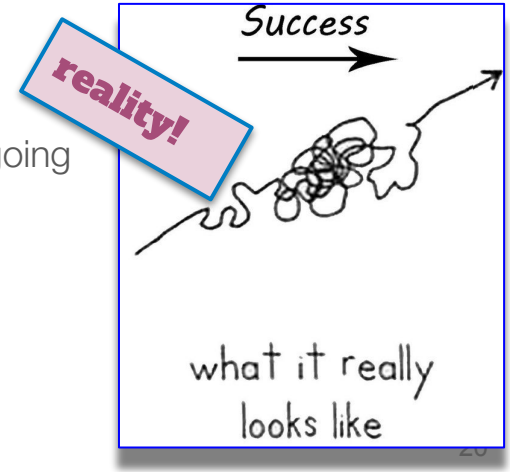


- Migration to AthenaMT *mostly* complete
- Getting ready for Run 2 Reprocessing (to bring data in line with Run 3)

Software areas status

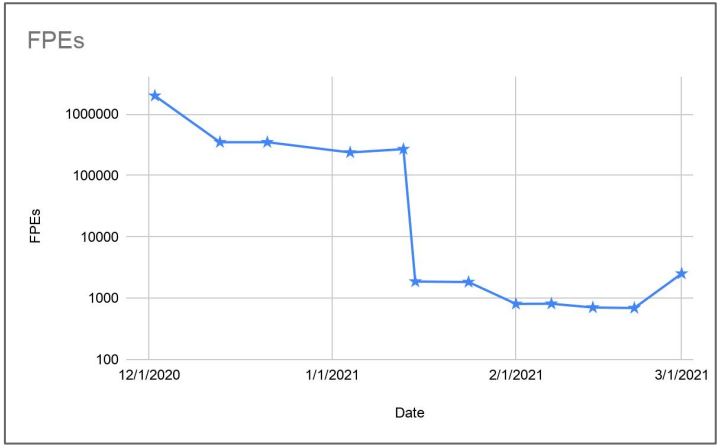
- Simulation
  - Geant4 Optimization TF bringing in many performance improvements
  - AF3 almost validated (5-8 times faster than Full Simulation)
  - Now integrating pre-mixed pileup (MC - MC overlay)
    - Exciting challenges on computing (RDOs usage wrt signal)
- Reconstruction
  - Most groups working on final tuning and improvements
  - A few remaining open issues which should be dealt in the next weeks/months
  - More details in next slide
- Analysis (Run 3 Analysis Model)
  - Most CP tools ported and tested in Release 22
  - DAOD_PHYS(LITE) will be the main format for Run 3(4)
    - DAOD_PHYS ~ 30-50kB/event (Data-MC)
  - Top level analysis framework migration to release 22 / DAOD_PHYS ongoing
  - Working with PA teams on extensive analysis-level Physics Validation
  - ~ 10 analysis already migrated to DAOD_PHYS

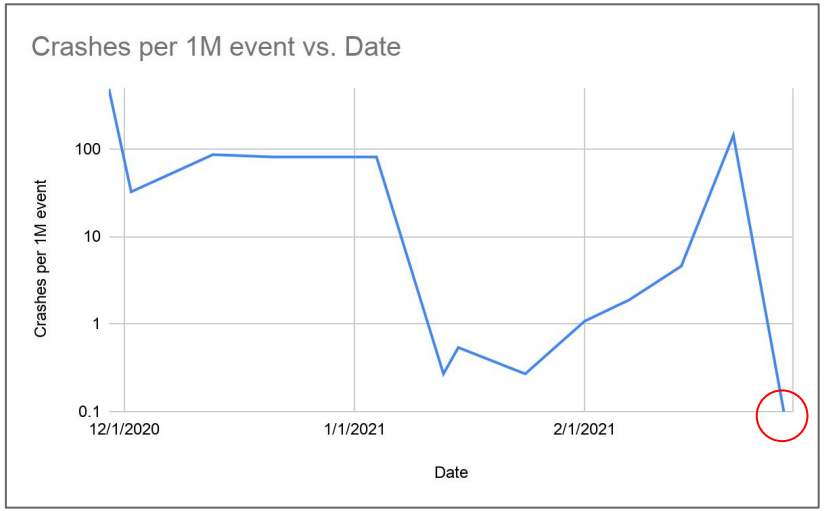




Release 22 reconstruction - more details

- In Dec 2020 started a campaign of ~weekly technical validation
  - Fixing one issue after another
- Readiness for Reprocessing criteria:
  - Per-event crash rate < 10^{-6} ; target < 10^{-8}
  - No irreproducibility
  - Successful Physics Validation





First time with zero Athena-related crashes in 3.7M data events from 2018

Activity planning

Data sample	Activity	2021 Q4	2022 Q1	2022 Q2	2022 Q3	2022 Q4
2022 data	Tier-0 reconstruction					
	Partial reprocessing					
	Full reprocessing					
	DAOD production and user analysis					
2022 MC	Simulation					
	Reconstruction					
	Re-reconstruction					
	DAOD production and user analysis					
2023 MC	Simulation					
Upgrade MC	Production and analysis					
2022 ions	Tier-0 reconstruction					

Run-3 readiness: AthenaMT performance

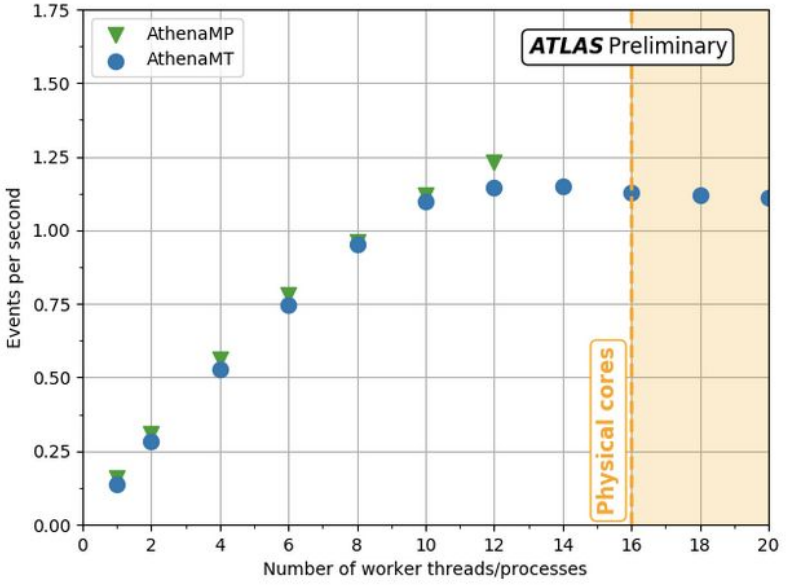


Figure 2 : Event throughput (events processed per second) as a function of number of threads/processes. The blue points show the multi-threaded Athena (AthenaMT) and the green points show the multi-process Athena (AthenaMP) results. The number of events processed concurrently is set to the number of threads in AthenaMT. Above 12 processes, AthenaMP exhausts the total usable memory on the node. Results are shown for the reconstruction of 500 simulated ttbar events including pile-up with an average number of interactions-per-bunch-crossing, <μ>, of 20. The orange dashed line shows the total number of physical cores on the node. The test is run on a machine with Intel(R) Xeon(R) CPU E5-2667 v2 @ 3.30GHz (16 cores/32 threads) and 32 GB of total memory.

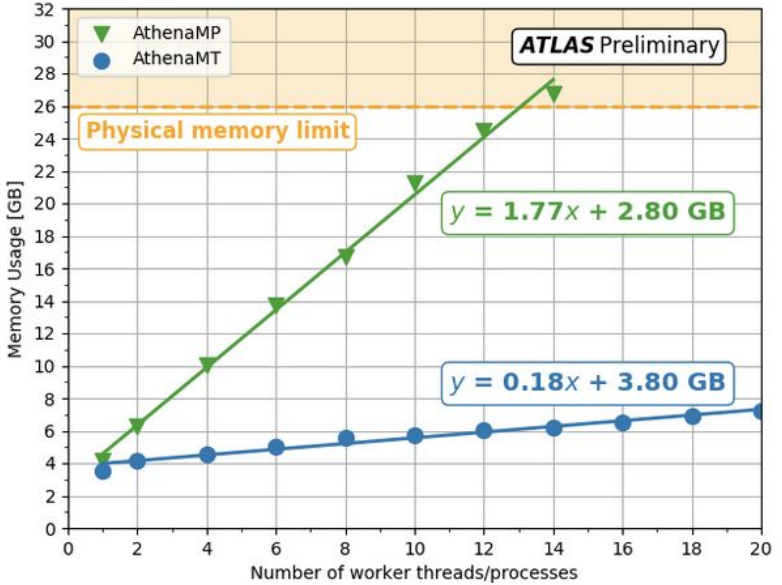


Figure 1 : Memory usage (measured by proportional set size) as a function of number of threads/processes. The blue points show the multi-threaded Athena (AthenaMT) and the green points show the multi-process Athena (AthenaMP) results. The solid lines are linear fits. The number of events processed concurrently is set to the number of threads in AthenaMT. Results are shown for the reconstruction of 500 simulated ttbar events including pile-up with an average number of interactions-per-bunch-crossing, <μ>, of 20. The test is run on a machine with Intel(R) Xeon(R) CPU E5-2667 v2 @ 3.30GHz (16 cores/32 threads) and 32 GB of total memory.

Run-3 readiness: Databases and Metadata

- Ongoing optimization of Conditions data model (COOL) and its distributed access in Run-3
  - Goal is to ease the migration to the new model (CREST) in Run-4
  - Simplify COOL folders and contents:
    - reduce channels and/or reduce IOVs; Pixel, MDT, Muon DCS
  - Review of using time series data (e.g. DCS) as Conditions:
    - data smoothing for some DCS folders; MDT, SCT, Tile
  - Cache length tuning to reduce distributed data access rate
- New ATLAS Databases And Metadata (ADAM) group to help longer-term coherent development for Run 4
- CREST developments ongoing (modern data model and better caching):
  - Reusing CMS ideas - fruitful collaboration
  - Server side and Client side implementations
  - Data migration tools from COOL to CREST
  - Redesign of the Conditions data access in Athena to read from COOL and/or CREST
  - Additional personpower available starting from January 2021

R2R4: Heterogeneous resources… a lot of things ongoing

- GPU on the Grid effectively running
  - ~ 40 GPUs, mainly for user analysis (ML), not yet fully exploited.
  - Not everything solved, not everything smooth.
- non-X86 machines (PowerPC and ARM)
  - Need access to a few builds + dev nodes
  - With LCG (and support)
- Notebooks for interactive usage are crucial
  - boost the integration within the experiment
  - share the expertise (also across experiments!)
- Workload (possible impact) still an open question
  - Chicken and egg issue: more use, more experience, more ideas.
  - paramount to build the expertise (e.g. tutorials?)
  - Bursty activities should be enabled
- Concerns:
  - Need of across-experiments (across-communities) shared efforts
    - → very good initiative Compute Accelerator Forum started 1st Oct
    - Is it enough?

GPUs on the Grid

GPUs in ATLAS

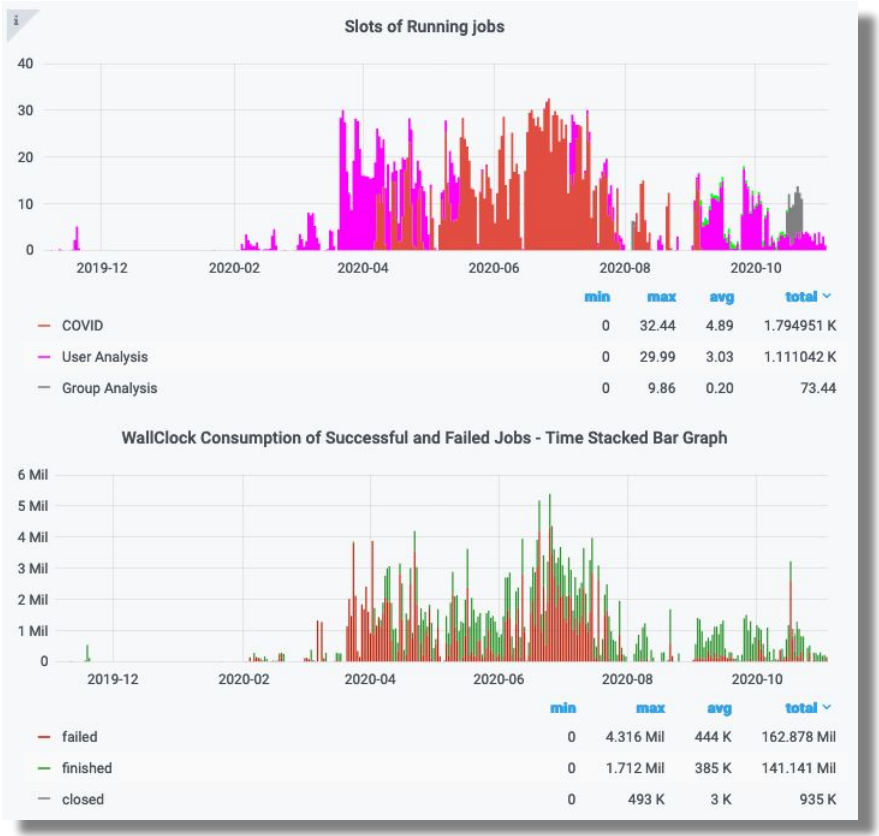
- 12 queues

Clone	ANALY_BNL_GPU_ARC
Clone	ANALY_CSCS-HPC_GPU
Clone	DESY-HH_GPU
Clone	ANALY_INFN-T1_GPU
Clone	ANALY_LRZ_GPU
Clone	ANALY_MWT2_GPU
Clone	ANALY_ORNL_Summit_GPU
Clone	ANALY_ORNL_Summit_GPU2
Clone	ANALY_OU_OSCER_GPU_TEST
Clone	ANALY_QMUL_GPU_TEST
Clone	ANALY_MANC_GPU_TEST
Clone	ANALY_SLAC_GPU

- Only 5 or 6 usable ATM
  - All grid sites
    - ~40 GPUs
  - Lot of work to get summit usable
- Problems
  - Opportunistic
    - User code is not check pointing
      - Even if it did resources may not reappear
  - No expectation of installed software at the site
    - particularly middleware
  - Memory not partitioned
    - Tensorflow didn't have limits set

8

[Compute Accelerator Forum](#) 1st Oct

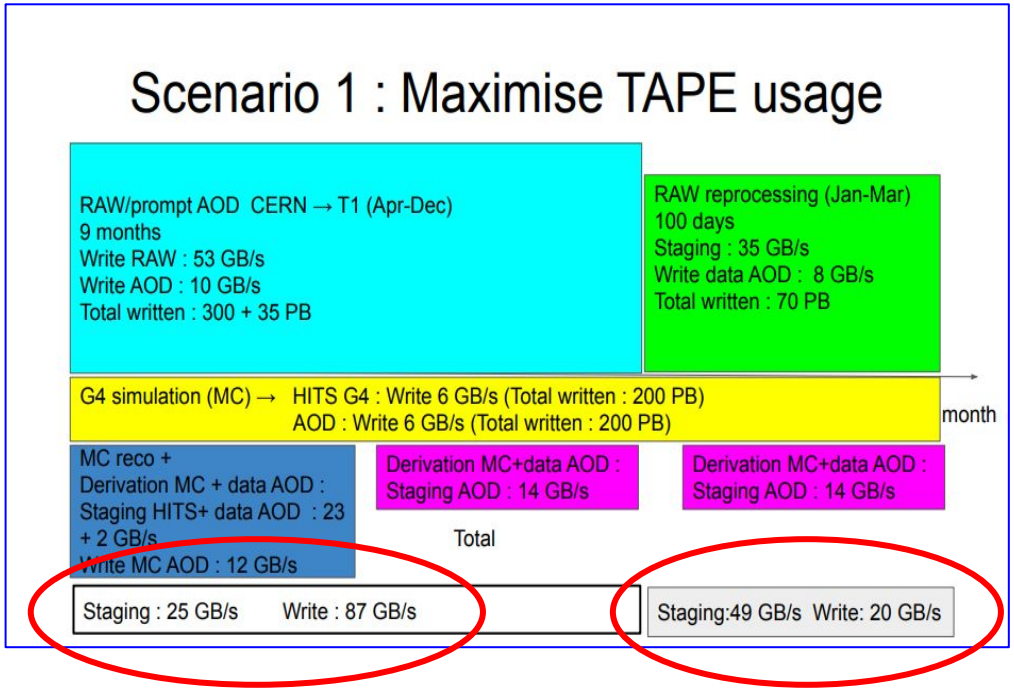


[ATLAS GPU on the Grid](#) (last year)

R2R4: T1s Tape challenges

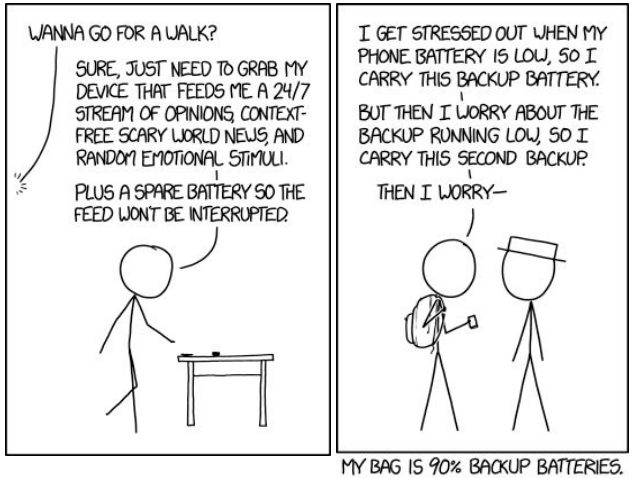
- Tape throughput for HL-LHC
  - First round of estimates
  - BNL: (23% of share)

Throughput	Jan-Mar	Apr-Dec
Read rate (delivered)	12 GB/s	6 GB/s
Read rate (nominal @ 50% eff.)	24 GB/s	12 GB/s
Write rate (delivered)	5 GB/s	20 GB/s
Write rate (nominal @ 80% eff.)	7 GB/s	25 GB/s



Atmosphere and brainpower

- Some reports of stressed people, too many meetings, etc (c.f. Feb GDB)...
  - Feeling that motivation is getting down
  - Difficult to engage new persons
- Working with WLCG to promote some actions that could help with this
  - Shorter, more focused general meetings
  - Make it clear *in advance* what summaries and outcomes are expected
  - Emphasis on 'quality interactions' (meeting attendance is a bad metric!)
- Increased complexity of getting people at CERN is a concern
  - Exacerbated by CERN rule changes
  - We hope this is resolved before Run 3 begins!



...

- ...
  - ...
- ...
- ..