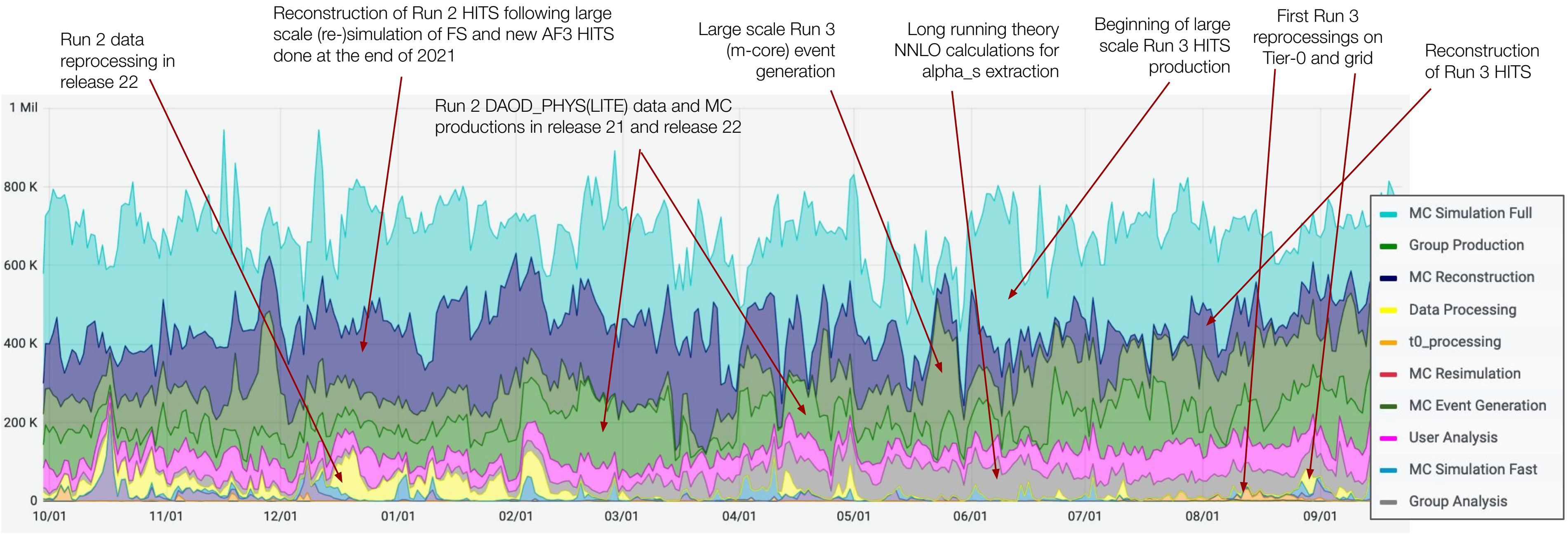
ATLAS Computing towards HL-LHC

WLCG Workshop
Lancaster University, 7th November 2022

David Cameron (University of Oslo), David South (DESY)

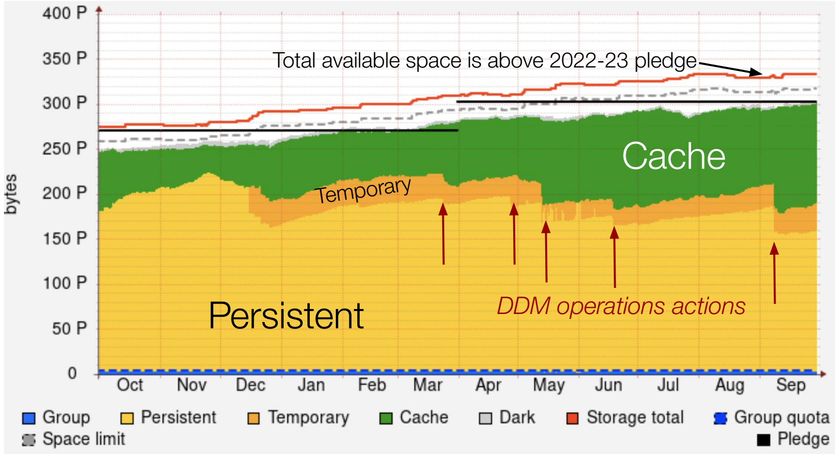
ATLAS running jobs in the last 12 months



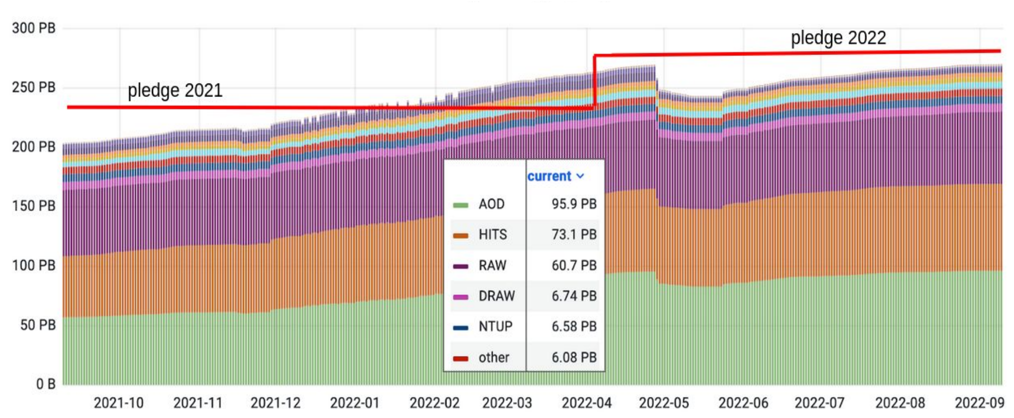
- Steady state of between 600k and 800k running job slots
  - Full spread of different activities, managed and prioritised via global shares
  - In addition to pledged Grid resources, significant contributions P1/HLT, Clouds and HPCs

ATLAS storage over the last 12 months

Total pledged disk



Total pledged tape

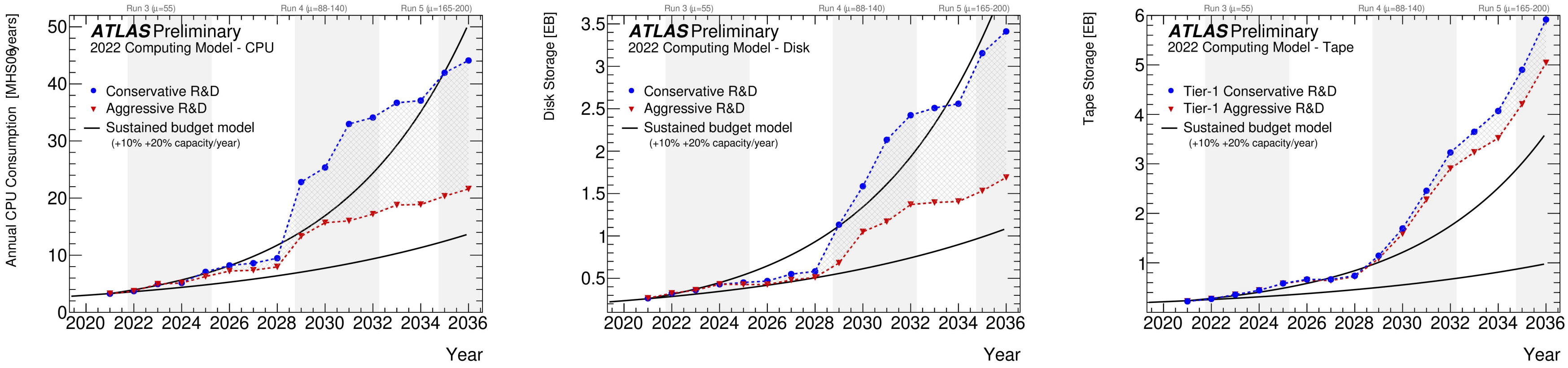


- Success of opportunistic compute comes at the price of constant storage pressure
  - Regular operations by DDM to keep it in check: removal of unused data, obsoletion of older versions,..
  - Use of data carousel in production workflows to ensure only what is necessary is on disk

- Likely to go over the tape pledge before the pledge year ends
  - T1 sites have responded positively to requests for early 2023 pledge deployment - much appreciated!

The HL-LHC computing challenge

- Projected evolution of resource usage from 2020 until 2036



Conservative (blue) and aggressive (red) R&D scenarios. The grey hatched shading between the red and blue lines illustrates the range of resources consumption if the aggressive scenario is only partially achieved. The black lines indicate the impact of sustained year-on-year budget increases, and improvements in new hardware, that together amount to a capacity increase of 10% (lower line) and 20% (upper line). The vertical shaded bands indicate periods during which ATLAS will be taking data.

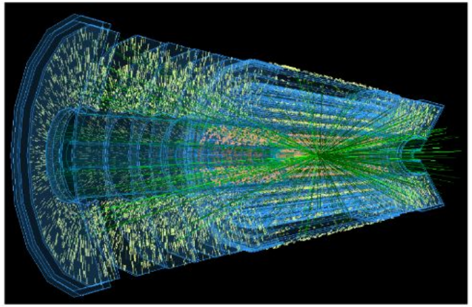
- We are all facing the requirement of a significant increase in computing resources for HL-LHC
  - Perhaps by working together, with some good people engaging in a few well defined, well thought out projects, we can save money in the future

The road to HL-LHC

- 2020: LHCC decided to perform a series of reviews of the Software and Computing plans of the LHC experiments towards HL-LHC
- The ATLAS HL-LHC Computing Conceptual Design Report was published in May 2020
- A follow up ATLAS Software and Computing HL-LHC Roadmap was published in March 2022 with concrete milestones
- A TDR is planned in ~2024

ATLAS Software and Computing

HL-LHC Roadmap



Reference:

Created: 1 October 2021
Last Modified: 22 February 2022
Prepared by: The ATLAS Collaboration

Milestones

- The roadmap defines a set of milestones per activity area:
  - Maintenance and Operations: essential just to get by
  - Conservative R&D: new developments achievable with current effort
  - Aggressive R&D: new developments requiring extra effort and/or successful development

Project Organization

MID	DID	Description	Due
PR-1		First evaluation of effort needed to deliver on HL-LHC milestones	Q2 2022
PR-2		HL-LHC Computing TDR	Q3 2024
	2.1	R&D projects targeting Run 4 ("Run 4 projects) define scope and potential impact of their demonstrators, and a program of work with effort and risk estimates to the end of Phase 2.	Q4 2022
	2.2	Define release, datasets and platforms to be used to evaluate Run 4 performance impact of demonstrators	Q1 2024
	2.3	Run 4 projects release their demonstrators	Q2 2024
	2.4	Run 4 projects evaluate the performance impact of their R&D demonstrators and estimate the effort needed to develop fully functional prototypes	Q2 2024
PR-3		Run 4 Release	Q2 2027
	3.1	Run 4 projects release fully functional prototypes, estimate risks and effort needed to bring to production quality	Q2 2026
	3.2	Run 4 developers tutorial	Q3 2026
	3.3	Run 4 Feature Freeze	Q2 2027
PR-4		Ready for Run 4 Data Taking	Q2 2029
	4.1	Run 4 projects demonstrate required functionality in release	Q3 2026
	4.2	Run 4 release validated	Q1 2029

Database infrastructure, conditions and metadata (ADAM)

MID	DID	Description	Due
DB-1		Relational database infrastructure consolidation	Q4 2022
	1.1	Migration of application schemas used in data processing from ATLR to ATONR	Q3 2022
	1.2	Setup of Frontier launchpad using ATONR_ADG (Active Data Guard) nodes.	Q4 2022
DB-2		Conditions database CREST development and integration	Q4 2026
	2.1	Define the CREST development, timeline and deployment strategy, as well as COOL data migration strategy.	Q2 2022
	2.2	Finalize the prototype for CREST deployment, partial COOL data migration and Athena jobs testing. Implement changes needed in IOVDbSvc.	Q4 2022
	2.3	Time series data processing and storage evaluation for CREST usage (DCS data).	Q1 2023
	2.4	Implement a directory structure to support selected conditions data from CREST into CVMS (access from sites without network connectivity).	Q4 2023
	2.5	Deployment of a parallel infrastructure for accessing conditions via CREST for larger tests (validate scalability and experts needs).	Q1 2026
	2.6	Final CREST validation for Run4 and COOL decommissioning.	Q4 2026
DB-3		Metadata evolution	Q1 2025
	3.1	AMI Application migration to CERN Oracle DB.	Q4 2021
	3.2	AMI Tag evolution and deployment.	Q4 2022
	3.3	Fully functional and integrated metadata system to characterize datasets (data and MC) used for physics analysis.	Q1 2025

Core Software, Heterogeneous Computing and Accelerators

MID	DID	Description	Due
CS-1		Pileup-digitization in AthenaMT production ready	Q4 2022
	1.1	Ensure reproducibility of MT production of presampled MB RDO files	Q2 2022
CS-2		Complete investigation of lossy compression techniques	Q4 2023
	2.1	Lossy compression of the ID track covariance matrix in the primary AODs	Q4 2021
	2.2	Lossy compression of DAOD	Q4 2021
	2.3	Lossy compression of primary AODs	Q4 2023
CS-3		Implement I/O roadmap metadata recommendations	Q4 2022
	3.1	Multi-threaded in-file metadata handling	Q2 2022
	3.2	Redesign of the metadata handling infrastructure (better support for fine-grained workflows)	Q4 2022
CS-4		Evaluation of data formats well-suited for massively parallel I/O (HPCs)	Q1 2022
	4.1	Storing intermediate EventService Simulation data in HDF5	Q1 2022
CS-5		Migration to ROOT 7	Q4 2026
	5.1	StorageSvc capable of writing a subset of DAOD_PHYS(lite) data to RNTuple	Q4 2023
	5.2	Migration of Athena from Root 6 to 7	Q1 2025
	5.3	ROOT and LCG release contributions, testing, feedback	Ongoing
CS-6		Re-evaluation of simulation data formats (including EVNT, HITS, RDO)	TBD
	6.1	Technical review of file storage format, compression, etc	
	6.2	Content review of data format	
CS-7		GPU Kernel scheduling	Q1 2023
	7.1	Basic support for Kernel scheduling in athena	Q2 2021
	7.2	Integration with Gaudi Scheduler	Q1 2023
CS-8		GPU management techniques and infrastructure in Athena	Q4 2021
CS-9		Develop Multi-algorithm heterogeneous applications	Q2 2023
	9.1	ACTS-based multi-algorithm workflow	Q3 2022
	9.2	FastCaloSim GPU merged into master	Q2 2022
	9.3	Calorimeter clustering	Q4 2022
	9.4	GPU-accelerated ML inference in athena	Q2 2023
CS-10		Infrastructure for processing data across multiple events on an accelerator	Q4 2023
	10.1	Proof-of-concept prototype	Q4 2021
CS-11		GPU Memory management	TBD
	11.1	First (Vecmem) prototype	Q1 2022
CS-12		Make ATLAS Data Model classes accelerator-friendly	Q4 2024
	12.1	Prototype GPU-friendly xAOD classes	Q1 2022
	12.2	Support for reduced/mixed precision in ATLAS EDM	Q4 2022
	12.3	Decision on xAOD API evolution	Q4 2022
	12.4	Event-batching and EDM	Q3 2023
	12.5	Accelerator-friendly detector data model (geometry and calibration)	Q4 2024
	12.6	Evaluate mechanism to offload and update detector description on device	Q4 2024
CS-13		Intra-node scheduling, targeting HPCs and grid	TBD
	13.1	Raythena/HPX-based scheduler prototype	Q3 2022

Simulation

MID	DID	Description	Due
SI-1		Updates required for MC+MC Overlay for Run4	Q2 2023
	1.1	Enhance Overlay code to better deal with high pile-up environments	Q2 2023
	1.2	Implement Overlay Algorithm for HGTD	Q2 2023
	1.3	Implement Overlay Algorithm for ITk Pixels	Q2 2023
	1.4	Implement Overlay Algorithm for ITk Strips	Q2 2023
SI-2		Track Overlay	Q2 2023
	2.1	Data overlay with pre-reconstructed tracks	Q2 2023
SI-3		Review or Run 4 Simulation Strategy	Q2 2024
	3.1	Evaluate Geant4 equivalents of ISF functionality	Q1 2023
	3.2	Evaluate accuracy of EMEC simulation in FastCaloSim	Q4 2023
	3.3	Prototype EMEC geometry using native Geant4-GPU constructs	Q4 2023
	3.4	Prototype GPU version of LAr sensitive detector callback	Q2 2024
SI-4		Pile-up Digitization in AthenaMT	Q4 2023
	4.1	Full AthenaMT compatible pile-up digitization	Q2 2023
	4.2	Reduce memory usage of pileup digitization for very high mu	Q4 2023
SI-5		Make data overlay useable for p-p collision simulation	Q2 2026
	5.1	Flexible alignment and geometry in simulation for data overlay	Q2 2024
	5.2	Skimming of zero-bias Bytestream data and efficient grouping of events as input to p-p overlay	Q2 2025
	5.3	Adapt Fast Chain workflow to data overlay	Q2 2026
SI-6		Fast simulation/FastChain development for Run4	Q2 2027
	6.1	R&D in FastChain alternatives	Q2 2026
	6.2	Development of substantial improvements to fast calorimeter simulation	Q2 2026
	6.3	Development of substantial improvements to fast inner detector simulation	Q2 2026
	6.4	Implementation of ITk in FATRAS	Q2 2026
	6.5	Tuning fast simulation to data	Q2 2027
	6.6	Improve Geant 4 interface for fast simulation	Q2 2026
SI-7		Geometry updates for the Run4 MC Campaign	Q4 2027
SI-8		Run4 optimisation of the performance of ATLAS full Geant4 simulation	Q4 2027
SI-9		Testing Geant4 Versions and configurations for Run4	Q4 2026
	9.1	Validate Geant 4 v11	Q4 2023
SI-10		Validation of full and fast simulation for Run4	Q3 2028
SI-11		Feature complete digitization + overlay for all sub-systems including conditions access	Q2 2027
SI-12		Feature complete trigger simulation including compatibility with overlay	Q2 2027

Reconstruction

MID	DID	Description	Due
RE-1		Merge detector upgrade release (21.9) with master	Q1 2022
RE-2		Adapt Reconstruction to the Phase-II Detector	Q1 2024
RE-3		Migrate default CPU chain to ACTS	Q3 2025
	3.01	Define and migrate to a new internal tracking EDM	Q1 2023
	3.02	Define run-4 ATLAS EDM	Q1 2022
	3.03	Migrate to the new run-4 ATLAS EDM	Q3 2022
	3.04	Full chain prototype/demonstrator with all the components	Q3 2024
	3.05	ITk reconstruction	Q4 2024
	3.06	Muon tracking geometry and navigation	Q3 2024
	3.07	Muon standalone reconstruction	Q1 2025
	3.08	Muon combined reconstruction	Q3 2025
	3.09	Calorimeter tracking geometry and navigation	Q3 2025
	3.10	Calorimeter track extensions	Q3 2025
	3.11	Electron and gamma reconstruction	Q3 2025
	3.12	Particle flow reconstruction	Q3 2025
	3.13	Tau reconstruction	Q3 2026
	3.14	Flavour tagging	Q3 2026
	3.15	Jets and Missing ET reconstruction	Q3 2026
RE-4		Accelerator and machine learning (R&D)	Q3 2025
	4.1	Develop demonstrators for accelerators and new ML techniques	Q1 2024
	4.2	Finalise and implement functional prototypes	Q3 2025
RE-5		Feature freeze	Q3 2026
	5.1	Calorimeter reconstruction	Q1 2026
	5.2	ITk reconstruction	Q1 2026
	5.3	Muon reconstruction	Q1 2026
	5.4	Electron and gamma reconstruction	Q1 2026
	5.5	Particle flow reconstruction	Q2 2026
	5.6	Tau reconstruction	Q2 2026
	5.7	Flavour tagging	Q3 2026
	5.8	Jets and Missing ET reconstruction	Q3 2026
RE-6		Performance freeze	Q1 2028
	6.1	Calorimeter reconstruction	Q2 2027
	6.2	ITk reconstruction	Q2 2027
	6.3	Muon reconstruction	Q3 2027
	6.4	Electron and gamma reconstruction	Q3 2027
	6.5	Particle flow reconstruction	Q4 2027
	6.6	Tau reconstruction	Q4 2027
	6.7	Flavour tagging	Q1 2028
	6.8	Jets and Missing ET reconstruction	Q1 2028
RE-7		Validation	Q3 2028

Analysis

MID	DID	Description	Due
AN-1		Baseline DAOD_PHYSLITE with run 3 s/w (event loop based) & TTree	Q3 2023
	1.1	First bulk production of prototype DAOD_PHYSLITE	Q4 2021
	1.2	Mechanism for evaluation of systematic uncertainties with PHYSLITE	Q2 2022
	1.3	Demonstrator for full analysis on PHYSLITE for target analyses	Q4 2022
	1.4	Recommendations for application of lossy compression	Q4 2022
	1.5	Finalised list of PHYSLITE contents	Q2 2023
	1.6	Development/roll-out of docs/training for run 3 analysis w/ PHYSLITE	Q3 2023
AN-2		PHYSLITE working with RNTuple	Q4 2023
	2.1	Implementation of RNTuple and revised xAOD in PHYSLITE	Q4 2023
AN-3		Prototyping & review of columnar data operations for end-to-end analysis	Q2 2024
	3.1	Tests of basic reading performance using TTree version of PHYSLITE	Q4 2022
	3.2	Prototyping of tools for columnar CP operations and other systematics	Q4 2023
	3.3	Adoption of ROOT7 data structures	Q1 2024
	3.4	Performance and ease-of-use assessment leading to decision on adoption	Q2 2024
AN-4		Development of columnar analysis infrastructure	Q2 2026
	4.1	Prototyping of framework for orchestrating columnar CP operations	Q3 2024
	4.2	Development of columnar skimming/augmentation	Q1 2025
	4.3	Demonstrate end-to-end analysis using columnar tools/fkw + dist. comp.	Q4 2025
	4.4	Development & roll-out of documentation/training for run 4 analysis	Q2 2026
AN-5		Accommodate all analyses in the run 4 analysis model	Q2 2028
	5.2	First assessment of run 4 analyses incompatible with PHYSLITE	Q2 2023
	5.3	Calculate resources implications and feed back to physics coordination	Q3 2023
	5.4	Set up DAODs for analyses which are unable to use the new model	Q2 2026
	5.5	Updated DAOD_PHYSLITE to capture additional analyses	Q2 2028

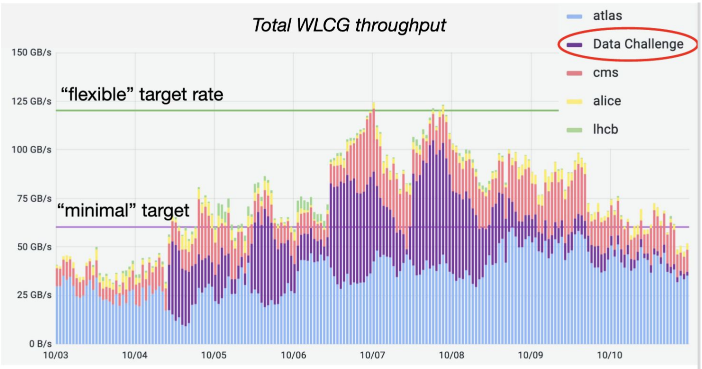
Distributed Computing

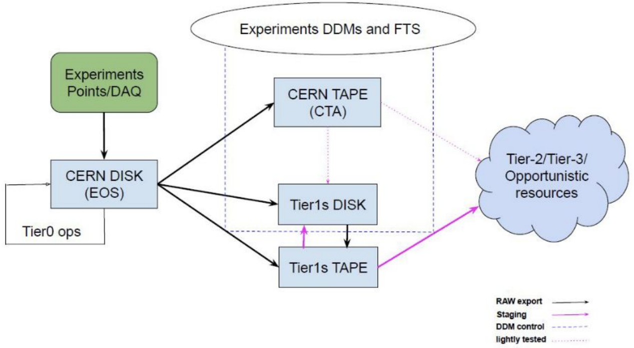
MID	DID	Description	Due
DC-1		Transition to tokens	Q4 2025
	1.1	Submission from Harvester to all HTCondor CEs with tokens	Q1 2022
	1.2	All users move from VOMS to IAM for X509	Q4 2022
	1.3	All job submission and data transfers use tokens	Q4 2025
DC-2		Storage evolution	Q4 2025
	2.1	No GridFTP transfers at any site	Q1 2022
	2.2	SRM-less access to tape	Q4 2025
	2.3	Recommended transition plan from DPM completed	Q4 2021
	2.4	Transition plan from all DPM sites	Q4 2022
	2.5	All sites moved away from DPM	Q2 2024
DC-3		Next operating system version	Q2 2024
	3.1	Ability to run on "future OS" on grid sites	Q4 2022
	3.2	Central services moved to "future OS"	Q4 2023
	3.3	(CentOS 7/8 EOL)	Q2 2024
DC-4		Network infrastructure ready for Run 4	Q4 2027
	4.1	Network challenge at 10% expected rate	Q4 2021
	4.2	Network challenge at 30% expected rate	Q4 2023
	4.3	Network challenge at 60% expected rate	Q4 2025
	4.4	Network challenge at 100% expected rate	Q4 2027
DC-5		Integrating next generation of HPCs	Q2 2023
	5.1	Integration of at least 2 EuroHPC sites	Q4 2022
	5.2	Integration of next generation US HPCs for production	Q2 2023
DC-6		Exploratory R&D on GPU-based workflows for next generation HPC	Q4 2023
DC-7		HL-LHC datasets replicas and versions management	Q2 2024
	7.1	Replicas and versions detailed accounting	Q4 2022
	7.2	DAOD replicas reduction	Q4 2023
	7.3	DAOD versions reduction	Q2 2024
DC-8		Data Carousel for storage optimization	Q4 2023
	8.1	Investigate with sites the cost of Tape infrastructure and the estimated cost in case of sensible increase of read/write throughput	Q4 2022
	8.2	Reduce the AOD on disk to 50% of the total AOD volume, using Data Carousel to orchestrate the stage from tape for DAOD production.	Q4 2023
DC-9		Disk management: secondary(cached) dataset	Q2 2023
	9.1	Evaluate the impact on job brokering and task duration if disk space for secondary data is reduced	Q2 2023

- Whoa - that's a lot of detail! Today I highlight the computing areas relevant to this meeting

Network data and tape challenges for HL-LHC

- A series of “Data Challenges” took place in 2021/22, testing the state of network and storage infrastructure
- Testing available bandwidth between main grid sites
  - 10% of HL-LHC bandwidth: 60 GB/s minimal; 120 GB/s flexible
    - Minimal and flexible targets achieved
  - Gradually increase transfer rate between now and Run 4: 30%, 50% and then finally 100% of the expected HL-LHC rate
    - This means *next step* is a **factor of three** compared to the previous data challenge (the biggest jump by far)
- Testing data read/write rates between CERN and T1 entities, as well as staging of data and transfer to T2s
  - Targets met for “Data Taking” (writes) and “After Data Taking” (also reading) modes of operation, as required for Run 3
    - Probably too early to look at real HL-LHC tape targets, considering required investment in new hardware





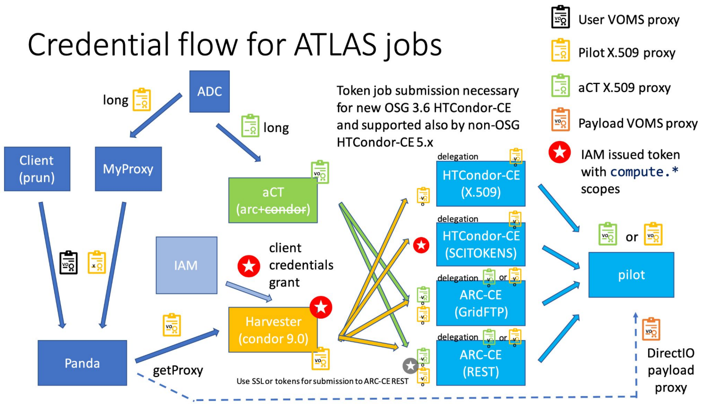
- Important to re-evaluate these targets and dates with respect the changes and/or delays in the HL-LHC schedule, as well as the experience gained during Run 3

Token transition

- An updated WLCG timeline was agreed upon in August
  - Feb 2023: At least one experiment moves away from VOMS (not necessarily ATLAS)
  - Feb/March 2024: Next HL-LHC Data Challenge using tokens
  - March 2026: Users no longer need X.509 certificates

- IAM support
  - Largely dependent on infrastructure support for tokens
  - Data management will be much more complicated than with X.509
  - Above timeline allows the request of extra effort from CERN IT to support IAM operations
  - Recent incidents with IAM job submission,
    whilst useful in the end for commissioning,
    illustrate **how crucial IAM support is**

- VOMS admins
  - Expect extra load during the transition from VOMS to IAM



HEPscore

- Two months ago there was an interesting workshop devoted to HEPscore, the possible successor to HS06 for compute node benchmarking
  - Composed of a number of LHC and other workloads running in containers
  - Several obvious advantages, including licensing issues
  - Next technical steps appear clear: finalize the definition of the benchmark suite

- ATLAS is broadly in favour of the following plan for moving to HEPscore for WLCG pledges
  - Provide a conversion factor for all old hardware from HS06 to HEPscore
  - Would then expect sites to benchmark all new purchases
  - We would be able to run the suite as well and externally measure the score of queues

  - 2025 pledge requests would be in both HEPscore and HS06 units
  - 2026 pledge requests would only be in HEPscore units

DPM EoL - and try to establish site set-up instructions?

ATLAS EXPERIMENT

- DPM retirement is set for June 2024 and several sites/federations already moved or are in transition
  - Various alternatives are being deployed, including EOS, CEPH, dCache..
    - Full details in EGI spreadsheet
  - Some sites will also move to storageless or integrate their disk pools with others
  - Expect a few sites which are late or never make it
    - Either accept the risk of unsupported storage or decommission them (likely mainly small sites)

- As there is no replacement “WLCG standard” storage for small sites, many different options are followed, making it hard to recommend a single solution to a new site
  - Sites opting for their own “non-WLCG standard” storage must provide long-term support to integrate it and address operational issues

- Is this indicative of something which could be improved?
  - ATLAS provides a set of recommendations to sites in terms of minimum configuration requirements
  - **Could WLCG go one step further and provide a set of instructions** on how to set up a site (e.g. deeply expand WLCGBaselineVersions and/or WLCGOpsWeb) as well as provide support (e.g. via WLCG discourse)

OS evolution

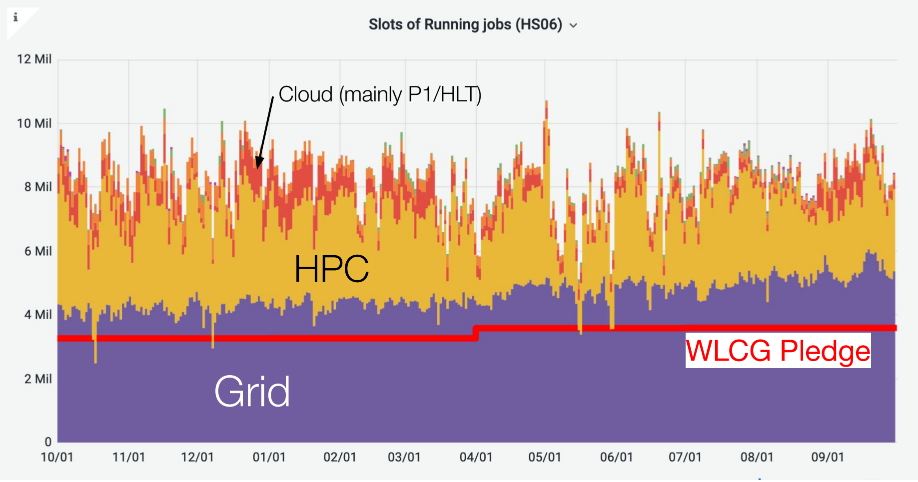
- Well known CentOS7 EoL is during Run 3, leading to search for the new OS
- From original proposal in December 21: Red Hat Enterprise, CentOS Stream and *Enterprise Linux Clones* should be treated equally
  - ELCs defined as 100% functionally equivalent to RHEL such as Rocky or Alma
- Until very recently it seemed the path was clear to CentOS (stream)
  - From Central Services perspective, plan was to skip CentOS8, and go straight to CentOS9
    - In fact first CS VMs running CentOS9 were deployed last week
  - From the grid perspective, as the payload is run in a container, worker nodes can be installed with CentOS8/9 and/or equivalent distros
    - Some worries, such as older ROOT versions embedded in ATLAS analysis software will stop being compatible with upgraded storage systems, but this is true for any new OS
- **ATLAS therefore notes with interest the discussion last week, namely that:**
  - There will be an extension of (partial) CentOS7 support until or even beyond the end of Run 3
  - CentOS9 stream may not be all it's cracked up to be..

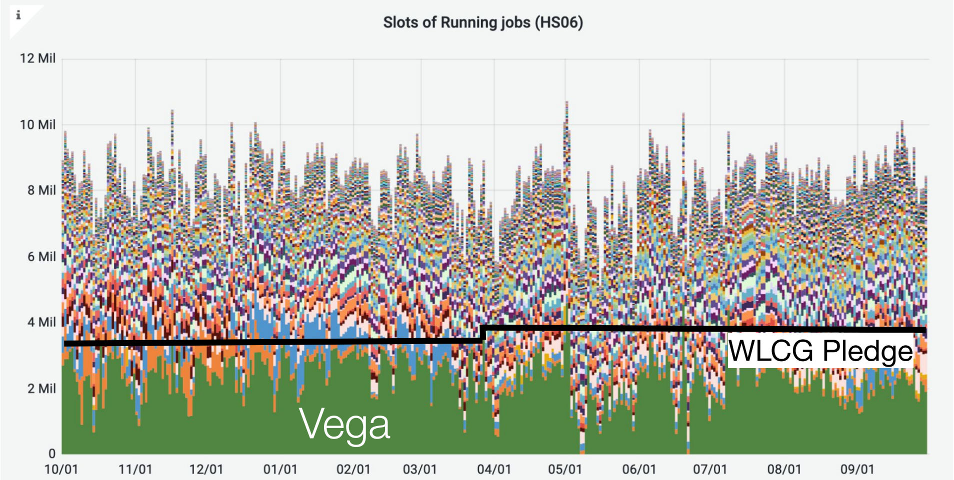
Resource evolution and impacts

- The Taiwan T1 transition to T2 In September 2023
  - Providing 25% of CPU and 50% of disk cf current T1+T2 pledge
  - Loss of ~1.5% CPU and disk
- Some funding agencies may consider changing their structure and/or how they provide pledged resources
- Rising electricity costs (CPU>>disk), delivery issues everywhere..
  - *More on this later*

- Around 2.5% of disk and CPU in Russia + JINR : **Continues for now**
  - CPU is fully used (above pledge)
  - Disks are only used to store transient data (no longer hosts unique data)
  - No new data to tape for last 2 years, but old data (MC only) is accessible. T1 is 2.1% total of pledge
  - NDGF network cut off from some sites (e.g Protvino)
- At the start of the crisis in February all unique data on RRC KI T1 disk was replicated elsewhere
  - Expectation was these resources could go away at any point, but up to now: business as usual
  - As things stand these sites will disappear at the end of 2024 (part of CERN decision)

Opportunistic resources: HPCs





- ATLAS continues to employ additional, non-Grid resources: HPCs, (commercial) cloud, HLT farm
  - HPC: *Vega*, *Karolina* (EuroHPC), several in US (*NERSC_Perlmutter* commissioning ongoing)
- Vega continues to deliver significant resources to ATLAS - CERN news article from the summer
  - Runs all workflows: our preferred set up, where the HPC is as close to grid site as possible
  - Flexibility managing production workload

- We cannot *rely* on such resources (e.g. Vega just reduced until mid-Dec): not comparable to Grid!
  - Pledged resources ***remain vital*** to ATLAS; however the level of opportunistic resources - and our exposure if they were to disappear - cannot be ignored. ATLAS will continue to employ HPC in the future

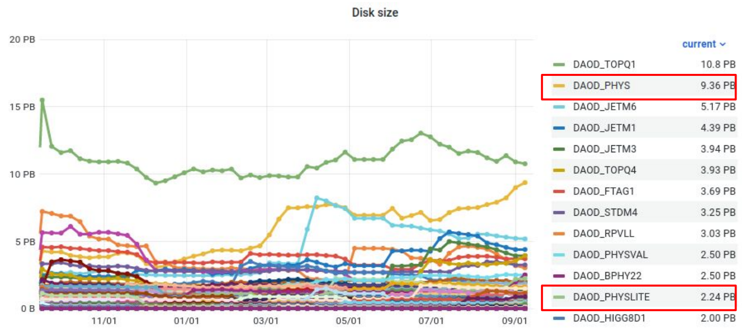
Infrastructure R&D: Additional and Opportunistic storage

- As shown earlier, for ATLAS the storage situation more critical than CPU
  - Happy to see pledges deployed early, both disk and tape
- How about additional tape at WLCG sites, in particular at the T2s?
  - There are T2s with substantial tape resources and expertise, and offers begin to emerge especially given the situation with Russia

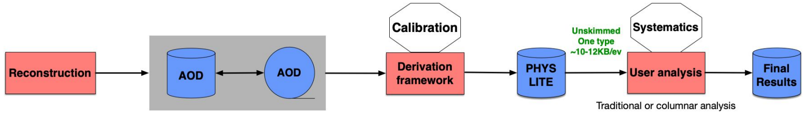
- We have experience with opportunistic compute - how about storage?
  - Non standard resources, transfer protocols, authentication/authorisation, network, firewall, ..
  - Most importantly, whilst compute can come and go, the data on the storage must remain accessible
- ATLAS is engaging with industrial partners on R&D projects
  - **SEAL**, a decentralised cloud storage start-up based on *FileCoin* - article
    - Some progress, simulation jobs reading EVENT running at AGLT2
  - Integration of **Google Compute Engine**, **Amazon Web Services**, **.. Lancium**
    - (Hopefully) only modest ATLAS expert time vs investment from partner
  - A Total Cost of Ownership of cloud resources is underway via the **ATLAS Google Project**
  - **Such projects provide interesting R&D opportunities and real resources, serving as risk mitigation**

Data analysis in Run 3 and HL-LHC

- Potential for large paradigm shift in HL-LHC driven both by expected resources crunch and explosion in new analysis techniques outside HEP
  - Small, common data formats with minimal information for (almost) all analyses
  - Python-based “columnar analysis” data science tools
  - Interactive rather than batch analysis



- Much of this is ready or being prototyped now
  - DAOD_PHYSLITE (compact common data format) originally planned for HL-LHC also being produced for Run 3 data and MC, as well as the larger DAOD_PHYS



- Run 3: Two DAOD production models in parallel for some time (in fact also AODs): Challenging!
  - Rel 21 Run 2 model with many DAODs ***and*** Rel 22 DAOD_PHYS(LITE), as well as “residual” DAODs
  - Recreation of DAOD on demand is a potential solution:
    - Instead of retaining old data via complicated user interaction, delete and recreate if necessary
    - One of several HL-LHC “demonstrators” being discussed

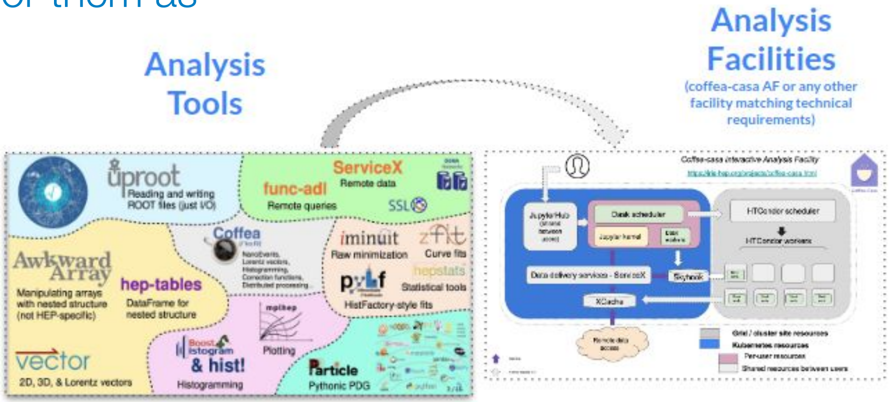
Analysis facilities

- A large amount of R&D (and hype) around analysis facilities.. .. which have a very broad definition
  - lxplus@CERN, NAF@DESY, this@there..
  - Jupyter with Dask on K8S behind federated AAI
- Can provide specialised hardware/software (GPUs, ML tools)
- Data delivery, caching and transformation services
- An unresolved question is if and how to account for them as part of pledged resources
- True and universal access also has issues
- Nevertheless, ATLAS follows developments in AF with interest, evaluating the evolution and assessing the implications

Requirements for AFs

- Modern authentication (AIM/OIDC), tokens, macaroons, scitokens
- Efficient data delivery and data management technologies
- Columnar analysis and support new pythonic ecosystem
- Modern deployment and integration techniques
- Support for object storage
- Efficient data caching solutions
- Easy integration with existing HPC resources

Alexander Held et al.



Heterogeneous and exotic architectures

- Another part of our HL-LHC milestones examines if we can exploit non-x86 architectures: GPU, ARM,..
- GPUs can potentially bring enormous benefits and are thus the highest priority R&D of the ATLAS core software team
  - Bulk processing
    - ML for tracking
    - Direct execution on GPUs
  - Analysis
    - ML training and inference
    - GPUs for standard analysis (e.g. fast statistics, histogramming)
- Connected to future infrastructure changes, e.g. exploitation of GPU-based HPCs, as well as deployment of ARM CPU
- **Note that all of these fall under aggressive R&D, hence may not be achieved without extra funding (and good fortune)**

RE-4		Accelerator and machine learning (R&D)	Q3 2025
	4.1	Develop demonstrators for accelerators and new ML techniques	Q1 2024
	4.2	Finalise and implement functional prototypes	Q3 2025

CS-7		GPU Kernel scheduling	Q1 2023
	7.1	Basic support for Kernel scheduling in athena	Q2 2021
	7.2	Integration with Gaudi Scheduler	Q1 2023
CS-8		GPU management techniques and infrastructure in Athena	Q4 2021
CS-9		Develop Multi-algorithm heterogeneous applications	Q2 2023
	9.1	ACTS-based multi-algorithm workflow	Q3 2022
	9.2	FastCaloSim GPU merged into master	Q2 2022
	9.3	Calorimeter clustering	Q4 2022
	9.4	GPU-accelerated ML inference in athena	Q2 2023
CS-10		Infrastructure for processing data across multiple events on an accelerator	Q4 2023
	10.1	Proof-of-concept prototype	Q4 2021
CS-11		GPU Memory management	TBD
	11.1	First (Vecmem) prototype	Q1 2022
CS-12		Make ATLAS Data Model classes accelerator-friendly	Q4 2024
	12.1	Prototype GPU-friendly xAOD classes	Q1 2022
	12.2	Support for reduced/mixed precision in ATLAS EDM	Q4 2022
	12.3	Decision on xAOD API evolution	Q4 2022
	12.4	Event-batching and EDM	Q3 2023
	12.5	Accelerator-friendly detector data model (geometry and calibration)	Q4 2024
	12.6	Evaluate mechanism to offload and update detector description on device	Q4 2024
CS-13		Intra-node scheduling, targeting HPCs and grid	TBD
	13.1	Raythena/HPX-based scheduler prototype	Q3 2022

CS-14		HL-LHC Technology decision: CUDA or one of its less-proprietary competitors	Q1 2024
	14.1	Full parallelization pattern recommendation to collaboration	Q1 2024
	14.2	Design patterns/tutorial on GPU migration	Q1 2024

Sustainability

- Some sites may need to reduce power consumption this winter
  - Hot topic given the current energy crisis, driven by the war in Ukraine
  - May come from site consumption target, price constraint, newly adopted law or even just voluntarily
  - ATLAS is actively looking at how to deal with the different scenarios if they occur

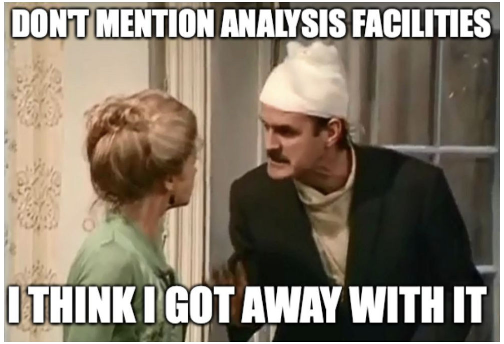
- If the number of sites becomes significant, may need coordination and preferential action
  1. *Turn off old hardware, during crisis or permanently - those with with lowest HS06/W*
  2. *Power down additional compute nodes: targeted saving - again those with highest W/HS06*
  3. *Compute cluster 100% powered down*
  4. *Storage disk nodes powered down - head nodes on, and regular turn on of all storage, not really viable*
  5. *Storage 100% down - worst case for ATLAS, requires forewarning and manual action for unique data*

- Without a successfully validated checkpointing solution, turning off a site for few hours a day to match peak load is not suitable for our computing model
  - Difficult to manage across various time zones?
  - Likely need to re-run jobs, having been killed after running for many hours: wasted walltime
  - Risk of hardware failure if it is constantly power cycled

- Lowering the frequency of the CPU to reduce power consumption may be a more viable option

Summary

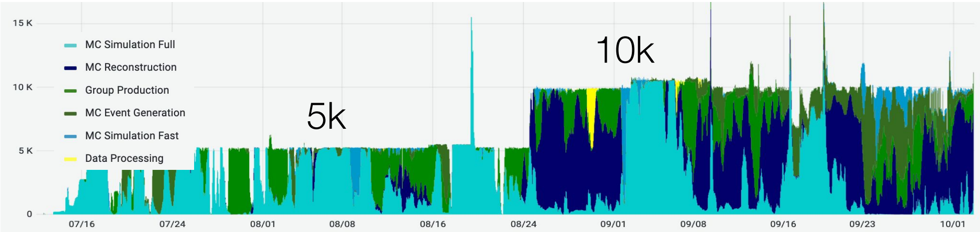
- ATLAS computing successfully runs on a diverse set of resources
- Run 3 has begun, adding another dimension to the diverse set of workflows
- Several operational changes are expected in the next period, which we follow closely
- We also look further ahead towards HL-LHC and the evolution in resources and analysis models
- Sustainability has quickly become a hot topic, and we are evaluating how this will affect us
- The majority of things presented here will be covered in more detail during the workshop - *we look forward to the discussions*



Back Up

Infrastructure R&D: ATLAS Google Project

- ATLAS Google Project (AGP):
  - 15 month collaboration between ATLAS and Google, begun this summer
  - Funding provided by US-ATLAS, but with exceptional rate and no risk of overspend
- Two main tracks:
  - An *R&D Track*, the continuation of the previous Google Cloud Project as already reported, with several new R&Ds now beginning
  - A *Cloud Site Track*, which integrates with "normal" ADC activities. This is in fact well progressed, and we have been running most workflows on Google since mid-July, with work ongoing now to include analysis



- In addition, an evaluation of the Total Cost of Ownership will also be performed
  - Several phases, starting by defining the metrics and culminating in a full TCO evaluation in late 2023
  - Establish the true cost of employing Cloud services for ATLAS workflows