

# ATLAS Computing update

WLCG referees meeting

10 September 2019

James Catmore, Davide Costanzo

Production update

Plans for 2021 and C-RSG requests

Update on HL-LHC strategy

## Smooth operation over the summer

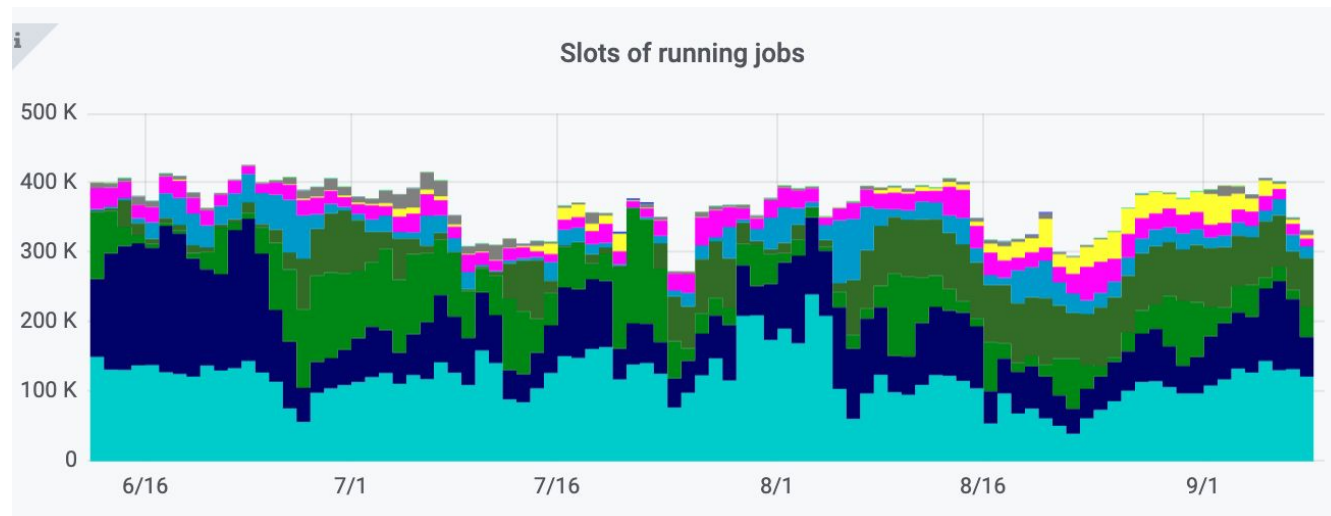
Successful production on Grid, Tier0, HLT (Mostly MC production to support Run-2 analysis)

Expect to operate in a very similar way for the next 18 months (Great thanks to our operations team!)

## Storage:

Applied lifetime model deletion on disk before the summer (and one planned soon)

Applied lifetime model deletion on tape last summer (so tapes can be repacked before Run-3)



## Smooth operation over the summer

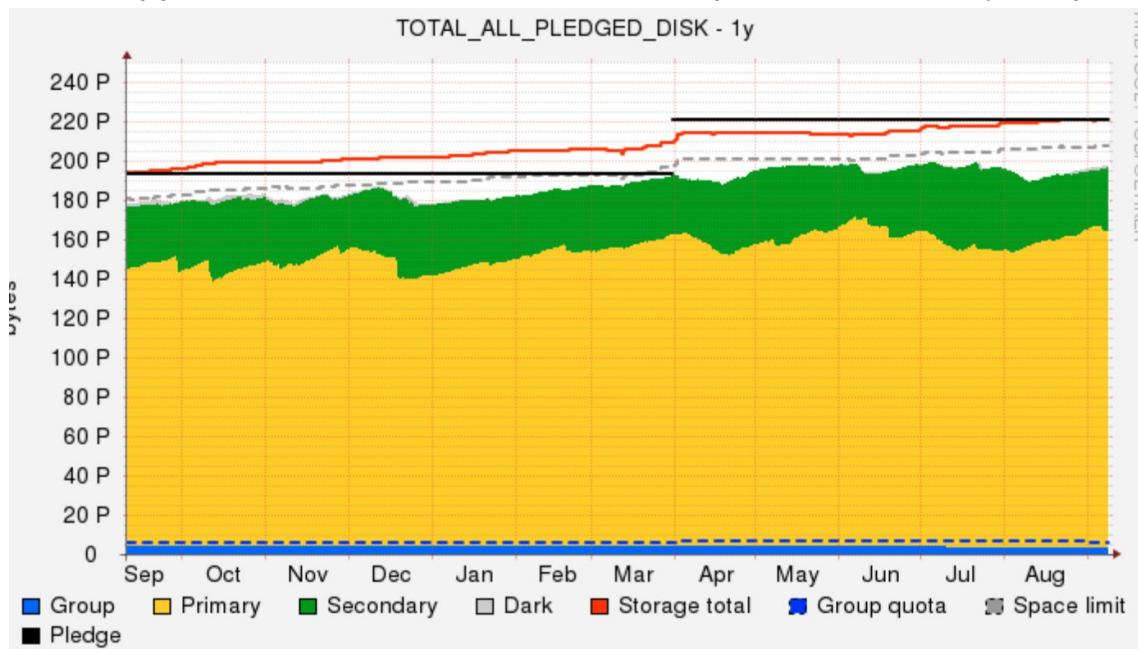
Successful production on Grid, Tier0, HLT (Mostly MC production to support Run-2 analysis)

Expect to operate in a very similar way for the next 18 months (Great thanks to our operations team!)

## Storage:

Applied lifetime model deletion on disk before the summer (and one planned soon)

Applied lifetime model deletion on tape last summer (so tapes can be repacked before Run-3)



Sample	Tier-0 reconstruction	Simulation	Grid reconstruction	DAOD production	
Run 2 data	All year		R22 reprocessing until summer	DAOD_PHYS(LITE) as AODs ready	
2021 data			DAOD_PHYS(LITE) CP+special DAODs as AODs ready		
Run 2 MC (MC16)			R22 reprocessing until summer	DAOD_PHYS(LITE) as AODs ready	
2021 MC			Starting January	Starting early 2021; Reprocessing October	DAOD_PHYS(LITE) CP+special DAODs as AODs ready
2022 MC			Starting October		
Upgrade and validation samples			All year	All year	All year

+ user analysis, heavy ions

Bunch intensity	$0 \rightarrow 1.4 \times 10^{11}$ protons/bunch, linearly rising over the year
Machine availability	<b>20%</b> (50% contingency)
Pile-up	Levelling at $\langle \mu \rangle = 55$ 4 hours of levelling per fill
Integrated luminosity	<b><math>17\text{fb}^{-1}</math></b> ( $42\text{fb}^{-1}$ contingency)
physics_Main events	<b>4 billion</b> (10 billion contingency)
MC simulation events	<b>20 billion</b> (for 2021+2022)

## Release 22:

- Multi-threaded software framework to improve memory utilisation and increase flexibility
- Reprocessing of Run 2 data and MC with R22. Uniform release for Run-2 and Run-3

## Fast simulation:

- FastCaloSimV2 will be used in Run 3: same CPU, better physics
- 50/50 ratio of fast/full simulation (to increase towards 75/25 during Run-3)

## Use of overlay for the simulation of pileup

## New analysis model:

- Significant reduction in number of derived AOD (DAOD) formats
- Physics groups will primarily use two new unskimmed formats with 30-50KB/event and 10KB/event respectively
- Only Combined Performance groups and special physics analyses to use existing DAODs

## Reductions in AOD size to be introduced during Run 3

Data carousel: on-demand recall of AODs from tape → less AOD needs to be kept on disk

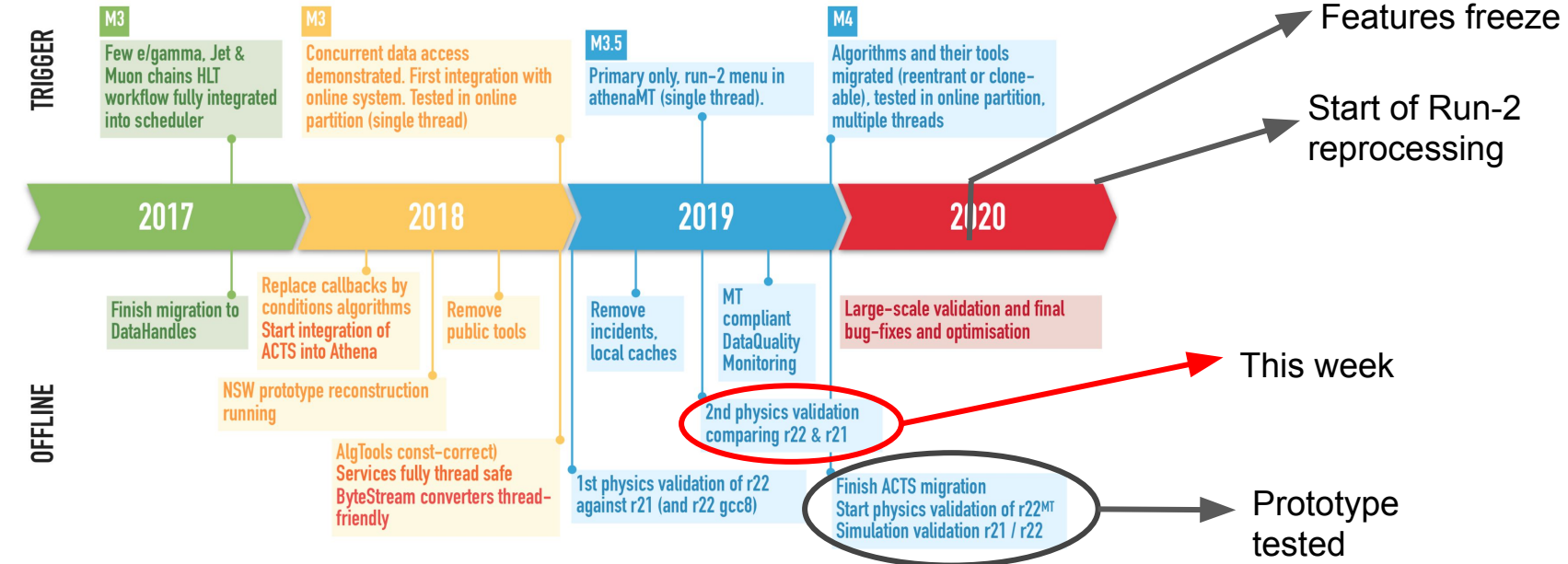
# Release 22 (aka Run-3 release) preparation

## Release 22: deployment of AthenaMT

- Currently running in multithreaded mode (offline and online) with one thread
- Multiple threads expected towards the end of the year
- Performance and optimisation in 2020

## Run-2 reprocessing planned with release 22 in 2020-21

Same release for Run-2 and Run-3 analysis



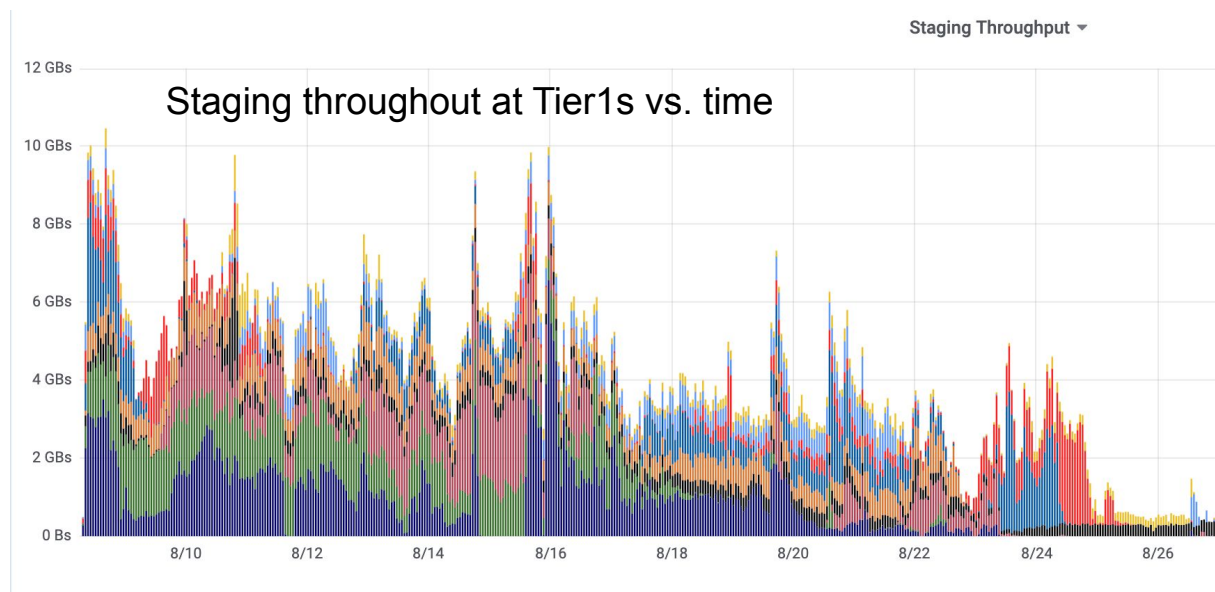
‘data carousel’ - an orchestration between workflow management (WFMS), data management (DDM/Rucio) and tape services whereby a bulk production campaign with its inputs resident on tape, is executed by staging and promptly processing a sliding window (10 % or more) of inputs onto buffer disk, such that only a small fraction of inputs are pinned on disk at any one time.

## Latest test in August

- Processing of 6.9 PB of data
- Started on 8 Aug, expected to last 2 weeks (5-6 GB/s)
- Some tails observed
- Technical challenges
- Discussion this week

## AOD access from disk is 100 PB/month

- For tape access we will have to schedule access in an orderly way





Improvement of Geant4 time/event

FastCaloSim: parameterisation of particle interactions and showers

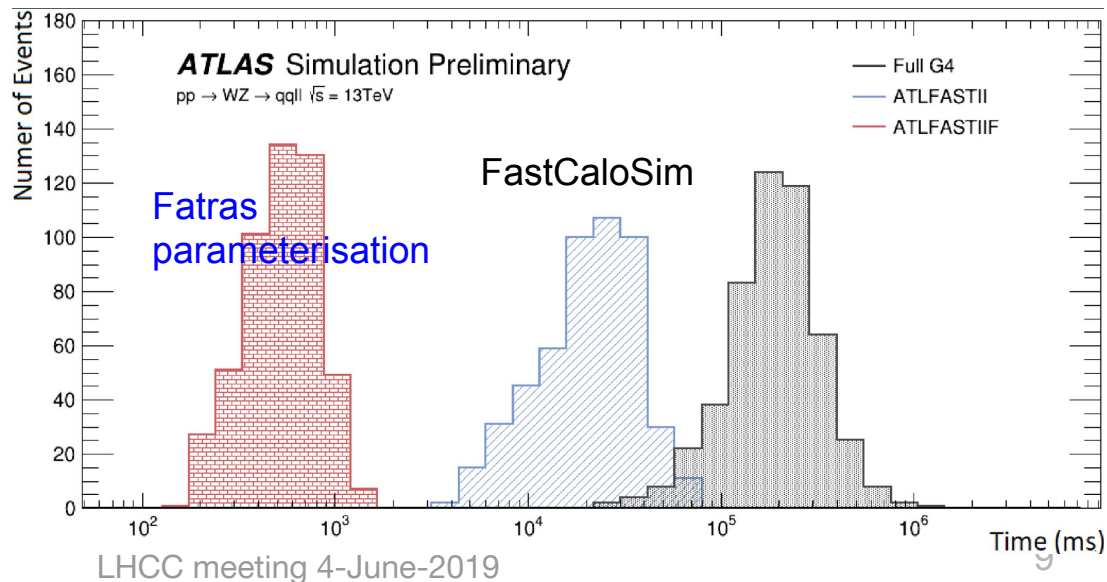
Further improvement possible with fast tracking (Run-4)

New version of Fast Calo Sim deployed for Run-3

Validation is crucial. Team of people working on it for the past 2 years.

Fast/full ratio to increase during Run-3

Current goal to have a 75/25 ratio at the end of Run-2

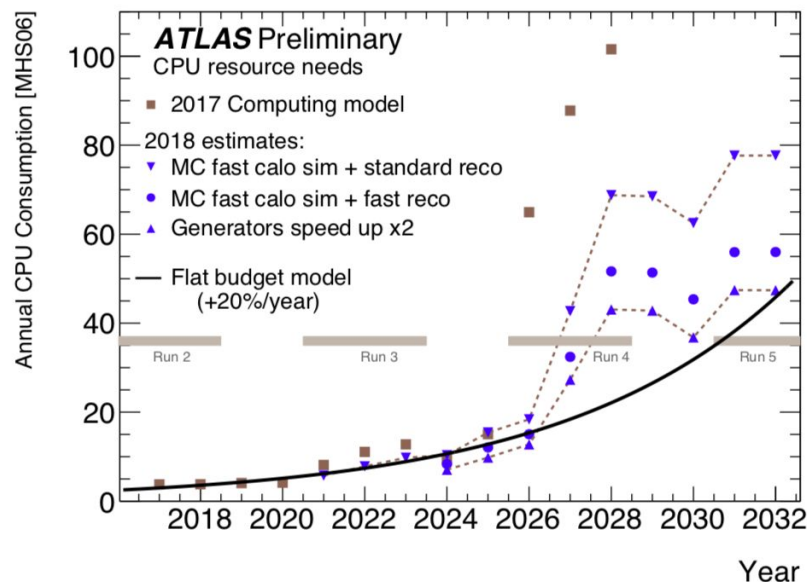
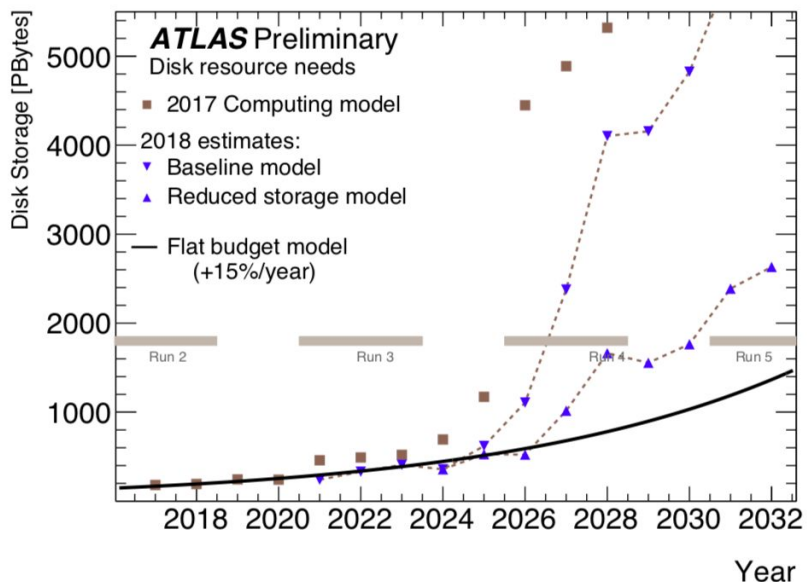


	2019 Agreed @ Oct 2018 RRB	2019 pledges	2020 Agreed @ April 2019 RRB	2021 Request @ Oct 2019 RRB	Balance 2021 wrt 2020 request
<b>T0 CPU (kHS06)</b>	496 (*)	496 (*)	496 (*)	550	11% (**)  <input type="checkbox"/>
<b>T1 CPU (kHS06)</b>	1057	1084	1057	1230	16%
<b>T2 CPU (kHS06)</b>	1292	1293	1292	1500	16%
<b>SUM CPU</b>	2760	2788	2845	3280	15%
<b>T0 DISK (PB)</b>	27	26	27	30	11%
<b>T1 DISK (PB)</b>	88	94	88	107	21%
<b>T2 DISK (PB)</b>	108	101	108	132	21%
<b>SUM DISK</b>	223	221	223	269	20%
<b>T0 TAPE (PB)</b>	94	94	94	97	3%
<b>T1 TAPE (PB)</b>	221	217	221	249	13%
<b>SUM TAPE</b>	315	311	315	346	10%

Update of the model (and plots) planned for Spring 2019  
Consistently with the LHCC review planning

Short document in preparation

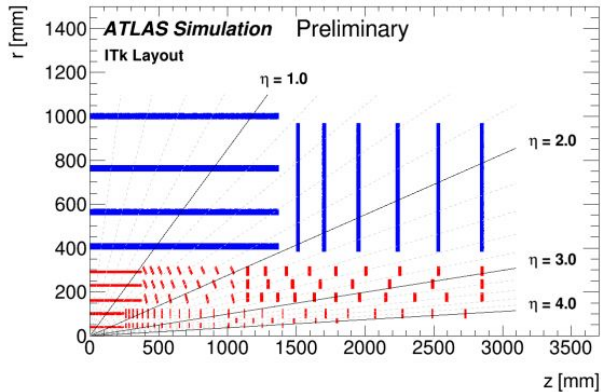
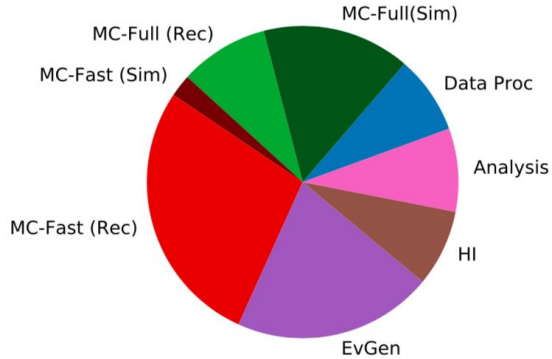
Several new idea in the past few months (in collaboration with HSF, WLCG, other experiments)



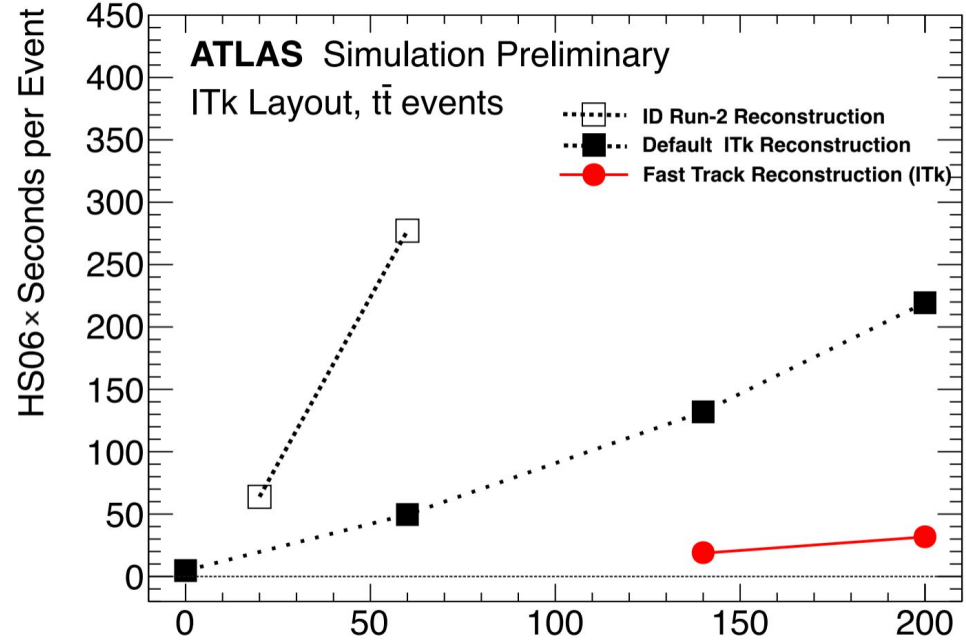
# Tracking reconstruction improvements

Need to speed up reconstruction at high  $\langle \mu \rangle$   
 Optimised tracker (ITk) with x10 more channels

ATLAS Preliminary. 2028 CPU resource needs  
 MC fast calo sim + standard reco



- Optimised track selection
- Improved seeding algorithm (for ITk)
- Omission of ambiguity resolving (to be partly recovered by the new fitter)



- Resources available to us on acceleration hardware (opportunistic)
  - Much debated topic in the past year or so
  - Complementary to non-x86 resources that could be used by recompiling the SW stack
- Prototype of Fast Simulation in CUDA
  - Self-contained kernel, collaboration with computing scientists
- ACTS module for GPUs in initial design phase (IRIS-HEP)
- Cross experiment initiatives:
  - Prospects of running event generators on GPUs.
  - Geant4 GPU kernels (?)
- Focus on frameworks for running on heterogenous resources
  - Two ATLAS senior developers charged with accelerators R&D
  - Current prototypes in CUDA running on NVIDIA GPGPUs.
    - Issue with sustainability, code duplication and validation
    - How practical is this outside Online or other contained environments?
    - How do we keep both CPU and GPU busy?
  - We need to focus on **portability**, kokkos, SYCL
  - Not all HEP code suitable for GPUs
  - The technology is evolving (We will soon evaluate Intel's OneAPI beta)

- Smooth operations during LS2
  - Fully using our resources for MC and analysis
  - Disk space monitored
  - Lifetime model deletions applied on disk and tape
- Resource requests for 2021 and Run-3
  - Request an increase of computing resources, after no increase in 2020
  - Several changes to the computing model (analysis model, tape carousel, fast simulation, ...)
  - New software release 22 with multithreaded framework
- HL-LHC computing
  - Progress on some R&D, tracking software, usage of GPUs and accelerators
  - Update and a document planned for Spring 2020 (following LHCC guidelines)
- Rotation in ATLAS Software and Computing management
  - Davide will retire on 1 October
  - James Catmore (coordinator) with Alessandro di Girolamo (deputy coordinator)