Improvements in ATLAS Computing for Run-3 of the LHC

Michal Svatoš

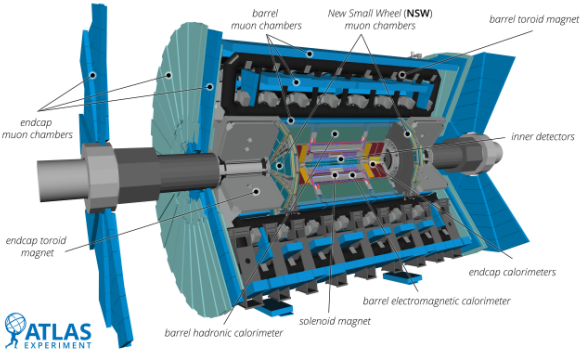
On behalf of the ATLAS Collaboration

ICHEP 2024, 18-24.7.2024

Introduction

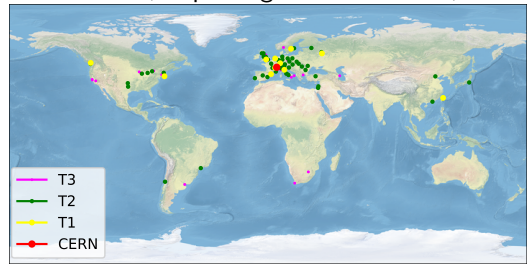
The ATLAS Experiment

- located at the Large Hadron Collider (LHC) at CERN near Geneva
- the collaboration comprises about 6000 members (about 3000 are authors of scientific papers)
- the detector is cylindric, 44m long, 25m in diameter, weighting 7,000 tonnes



The ATLAS Distributed Computing (ADC)

- manages resources at more than 250 sites located around the world, which provide nearly one exabyte of storage capacity and close to one million of CPU cores
- runs 24 hours/day, 365 days/year
- structure is evolving to give more flexibility to people, adapting to how they contribute, improving communication, etc.



Software

Core software

- Athena
  - offline software framework
  - open source (Apache 2.0 License)
  - mostly C++, configured using Python and built using CMake
- multiple external dependencies, including ROOT

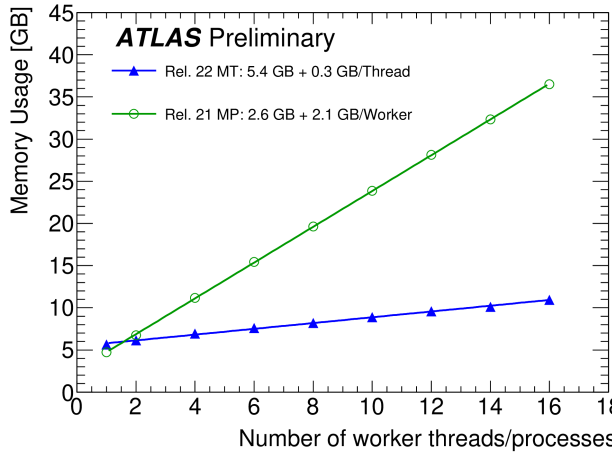
Language	Files	Comment	Code
C++	17,273	457,373	2,608,231
Python	9,478	211,655	1,009,088
C/C++ Header	20,475	469,490	843,679
Custom Configuration	307	0	368,828
XML	954	12,800	204,169
Shell	1,243	12,283	48,782
CMake / make	2,070	11,021	35,751
Fortran	166	7,674	24,024
Web (HTML, CSS, PHP)	44	289	7,085
CUDA	28	648	5,445
Other	171	3,235	24,027
Total	52,191	1,186,288	5,178,472

Number of files, comment lines, and code lines in the Athena software repository, divided by programming language (a snapshot).

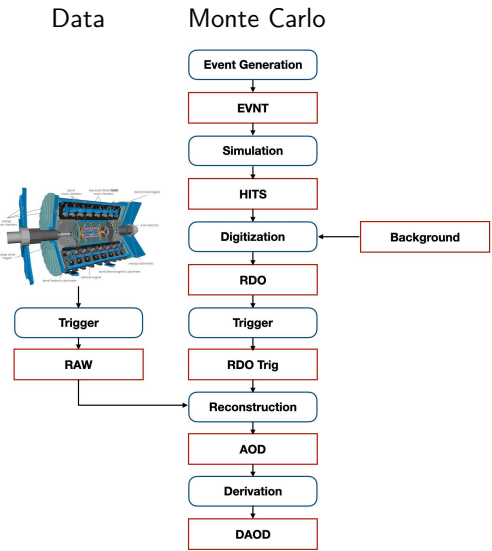
Software

Examples of improvements in Athena

- multi-threading
  - multiple events may be processed concurrently
  - significantly reduces memory requirements
- configuration
  - python based
  - previous system (job options) was difficult to maintain, extend, and debug
  - current system (Component Accumulator) is more modular



Workflows



Monte Carlo

- event generation - generation of the event record from a theoretical framework
- simulation - propagation through the material of the ATLAS detector of all generated particles which escape the beam pipe
- digitization - emulation of the detector response
- trigger simulation - a menu of triggers specific to the MC process is emulated and decisions and key trigger object collections are added to the output
- reconstruction - provides analysis object data (AOD) outputs, representing a summary of the reconstructed event
- derivation - produces intermediate-sized data types for analysis

Data

- data produced by the detector are reconstructed and turned into derivations

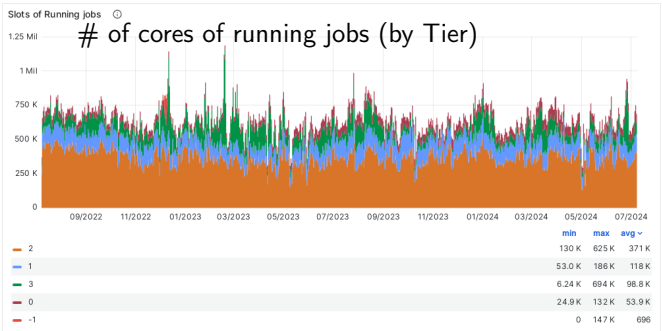
Workflows

Examples of improvements in workflows

- event generation
  - usage of HepMC3 event record in event generation (enables event numbers with up to 64-bit precision)
- full simulation
  - improvements in performance (outcome of task force with the mandate of achieving > 30% CPU performance speedup), e.g. usage of VecGeom, GEANT4 static linking, etc.
  - improvements in modelling (Birks' law tuning, Woodcock tracking, etc.)
- fast simulation
  - AtlFast3 (replacing the calorimeter shower simulation with faster methods)
  - FastChain (combining AtlFast3 with fast tracker simulation and track overlay for pileup simulation)
- digitization
  - MC overlay (digitised hard-scatter events are then combined with pre-digitised pile-up events) reduces CPU, memory and I/O requirements
- reconstruction
  - many small changes in tracking and calorimeters
- derivations
  - introduction of DAOD_PHYS and DAOD_PHYSLITE

Distributed Computing

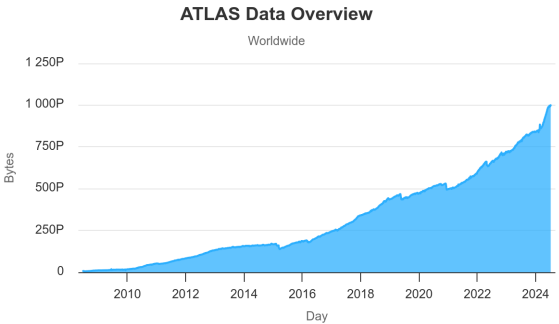
- Distributed Data Management (DDM)
- Workflow Management System (WFMS)
- computing centers
  - Tier-0 (CERN)
    - hardware, software and operations needed to support the prompt processing of the data produced by the ATLAS detector
  - Tier-1 (11 sites)
    - pledge CPU, disk, and tape storage and are charged with the perpetual archival of detector data
  - Tier-2 (∼70 sites)
    - pledge CPU and disk storage
  - Tier-3
    - no pledged resources



Distributed Computing: Distributed Data Management

Distributed Data Management = rucio

- Rucio is a flexible and modular software framework to build data management federations
- capable of high interaction rate
  - for example, in 2022 it was typically beyond 200 Hz and often reached 500 Hz using single Rucio instance at CERN



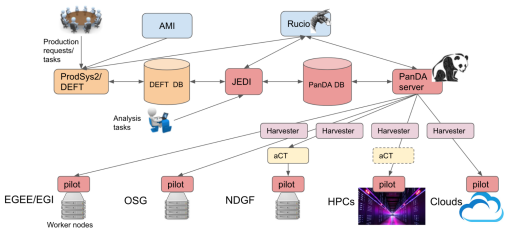
Examples of improvements

- **Data Carousel**: allows more active use of tape
- **security**: ongoing migration from X509 certificates to tokens from WLCG Authentication and Authorisation Infrastructure (AAI)
- **integration of commercial clouds**: Google, Amazon, Seal
- **integration of caching services**: XCache

Distributed Computing: Workflow Management

Workflow Management = ProdSys2-PanDA

- orchestrates processing, simulation and analysis of data



Examples of improvements

- integration of new cloud/HPC resources
- **iDDS**: developed to optimise resource usage in various workflows
- **Harvester**: enables more intelligent workload management and dynamic resource provisioning
- **security**: ongoing migration from X509 certificates to tokens from WLCG Authentication and Authorisation Infrastructure (AAI)

Databases

- Oracle database infrastructure (maintained by the CERN IT department)
  - rucio, PanDA
  - databases provide non-event data (the run configuration, detector calibration and alignment, etc.)
- Hadoop; Elasticsearch/Opensearch
  - analytics

Improvements

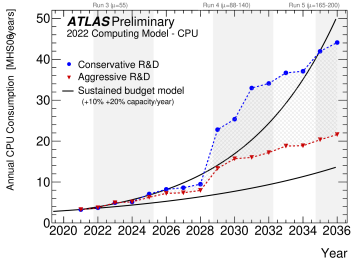
- consolidation of Oracle databases in CERN to decrease license/replication cost

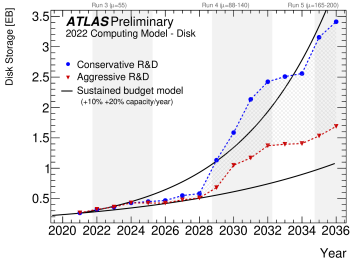
preparation for HL-LHC

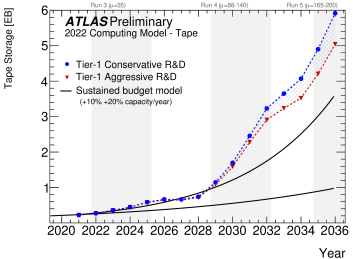
- WLCG Data Challenges
  - increasingly challenging stress tests of the data infrastructure, including disk, network, and the tape systems
  - The plan is to run these tests every two years until 100% of the expected HL-LHC data rate is reached
  - they are successfully revealing bottlenecks within middleware, infrastructure, etc.
- analysis
  - analysis facilities (dedicated infrastructure for analysis)
  - possibility to use columnar analysis
- usage of non-x86 architectures (ARM processors) and GPUs

More in **ATLAS Software and Computing HL-LHC Roadmap**
(`https://cds.cern.ch/record/2802918`)

Summary and Conclusions







- a lot of R&D is needed to fit into available resources (especially for HL-HLC with an order-of-magnitude increase in data volume)
- significant effort is needed to maintain the system while it is being improved and optimized
- in preparation for the HL-LHC, a 10-year plan was laid out and its milestones are followed

Run3 paper

Software and computing for Run 3 of the ATLAS experiment at the LHC
https://arxiv.org/pdf/2404.06335