Data Processing in ATLAS

Scientific Computing Session

Wolfgang Walkowiak

Universität Siegen

HEP Experimental High Energy Physics
CPPS Center for Particle Physics Siegen

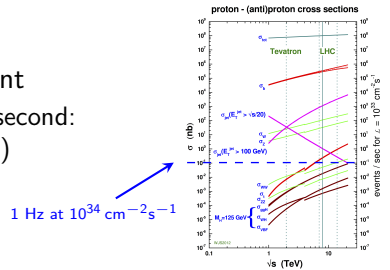
CPPS Retreat 2024, Meinerzhagen

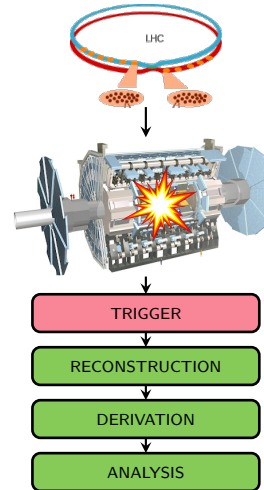
February 15, 2024

Bundesministerium für Bildung und Forschung

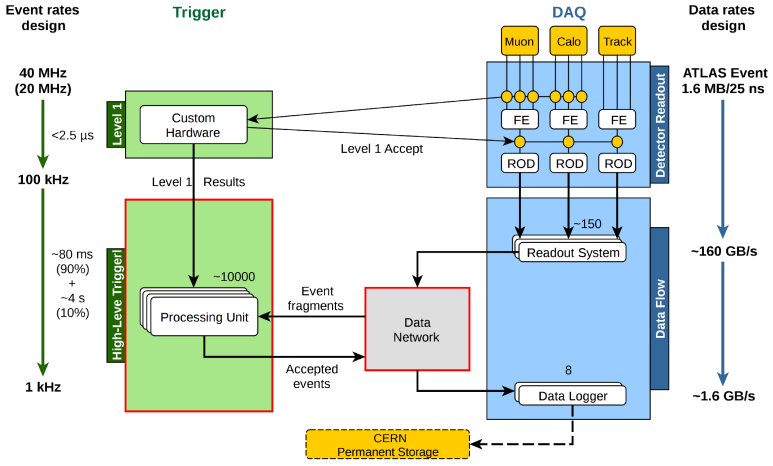
ATLAS Data Taking

- LHC collision rate: 40 MHz
- ATLAS produces ~ 1.8 MB/event
  - At $\mathcal{L} = 2{\times}10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ per second:
    $\sim 20\,000$ dijets ($p_{\mathrm{T}} > 100$ GeV)
    ~ 2000 W bosons
    ~ 1000 Z bosons
    ~ 50 top quarks
    $\sim 1/2$ of a Higgs boson
  - $\Rightarrow$ 80 TB/s impossible $\rightarrow$ 8 GB/s ok
- Trigger system:
  - reduce rate
  - keep balanced selection prioritizing *interesting* events
- Event reconstruction on Grid
- Easy parallelization: events are independent





ATLAS Online Trigger & Data Acquisition System



[graphics by J. Catmore]

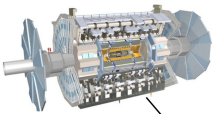
Hardware trigger (LVL1)

- Decision in $\sim 2.5\,\mu$s
- Retains $\sim 100\,$kHz

High level trigger (HLT)

- Offline-like algorithms
- Decision in $\sim 500\,$ms
- Max. 8 GB/s output
  - $\sim 3\,$kHz of full events
  - $\sim 5\,$kHz of 'partial' events

ATLAS Event Processing Pipeline & Monte Carlo



Data

GENERATION

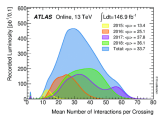
SIMULATION

DIGITIZATION

RECONSTRUCTION

DERIVATION

Monte Carlo



Pileup overlay
$\sim 30 - 60$ collision/event

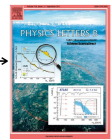
TRIGGER

RECONSTRUCTION

DERIVATION

ANALYSIS

Online
Tier-0
Grid
Local



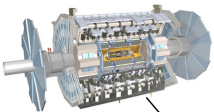
[graphics by J. Catmore]

MC generators

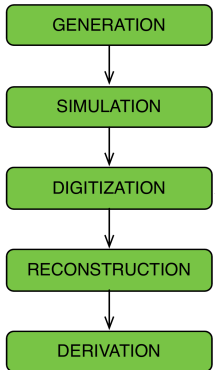
- MadGraph, POWHEG-BOX, Pythia8, Herwig7, Sherpa, ...
- MC production requests!

$\rightarrow$ 1 s to n hours/event

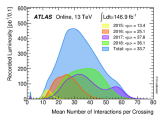
ATLAS Event Processing Pipeline & Monte Carlo



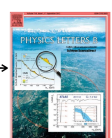
Data



Monte Carlo



Pileup overlay
$\sim 30 - 60$ collision/event



[graphics by J. Catmore]

MC generators

- MadGraph, POWHEG-BOX, Pythia8, Herwig7, Sherpa, ...
- MC production requests!

$\rightarrow$ 1 s to n hours/event

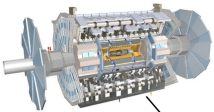
Detector simulation

- Mainly GEANT4
- Esp. CPU intensive: calorimeter simulation

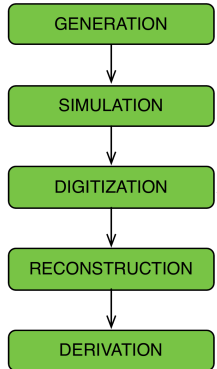
$\Rightarrow$ FullSim often replaced by AtlFast

$\rightarrow$ 1 to 10 min/event

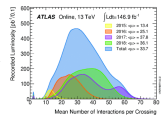
ATLAS Event Processing Pipeline & Monte Carlo



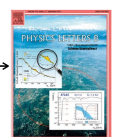
Data



Monte Carlo



Pileup overlay
$\sim 30 - 60$ collision/event



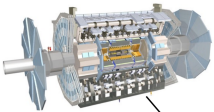
[graphics by J. Catmore]

Digitization

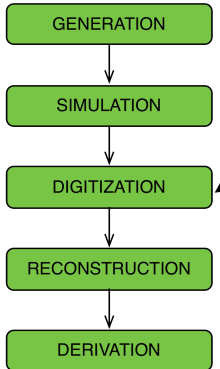
- Material interactions
  $\rightarrow$ realistic signals
- Overlay additional minimum bias pp events (*pileup*)

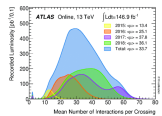
$\rightarrow$ 5 to 60 s/event

ATLAS Event Processing Pipeline & Monte Carlo

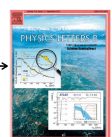


Data





Pileup overlay
$\sim 30 - 60$ collision/event



Online
Tier-0
Grid
Local

[graphics by J. Catmore]

Digitization

- Material interactions
  $\rightarrow$ realistic signals
- Overlay additional minimum bias pp events (*pileup*)

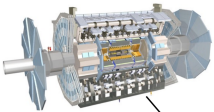
$\rightarrow$ 5 to 60 s/event

Event reconstruction

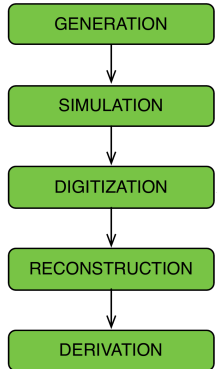
- Detector signals
  $\rightarrow$ physics objects
- Same algorithms as for real data
- Produces AODs

$\rightarrow \mathcal{O}(1\,\text{min})$

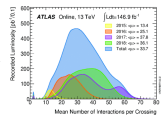
ATLAS Event Processing Pipeline & Monte Carlo



Data

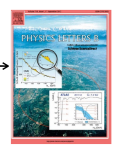


Monte Carlo



Pileup overlay
$\sim 30 - 60$ collision/event

[graphics by J. Catmore]



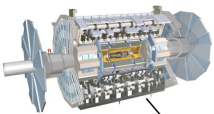
Derivation

- Reduction of AOD content for analysis
- Construction of analysis-specific objects (e.g. B-mesons)

$\rightarrow$ many formats (number to be reduced)

$\rightarrow$ $\mathcal{O}$(seconds/event)

ATLAS Event Processing Pipeline & Monte Carlo



Data

GENERATION

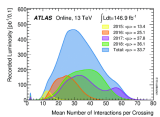
SIMULATION

DIGITIZATION

RECONSTRUCTION

DERIVATION

Monte Carlo



Pileup overlay
$\sim 30 - 60$ collision/event

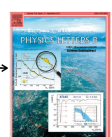
TRIGGER

RECONSTRUCTION

DERIVATION

ANALYSIS

Online
Tier-0
Grid
Local



[graphics by J. Catmore]

Derivation

- Reduction of AOD content for analysis
- Construction of analysis-specific objects (e.g. B-mesons)

$\rightarrow$ many formats (number to be reduced)

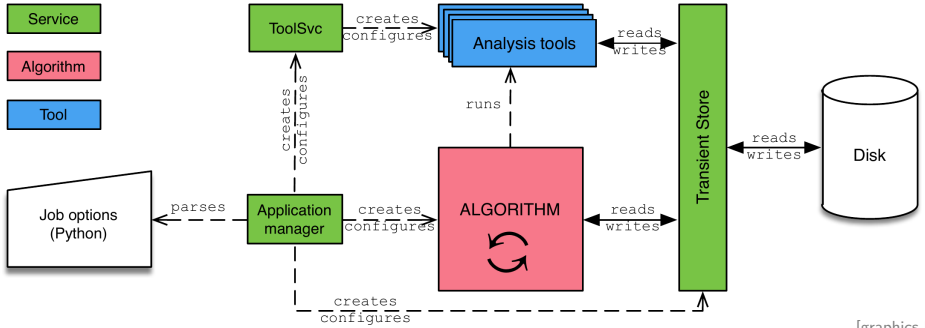
$\rightarrow$ $\mathcal{O}$(seconds/event)

Analysis

- Ntuples or data frames
- S/W specific to analysis

$\rightarrow$ takes months to years (physicists at work ...)

Athena Software Framework

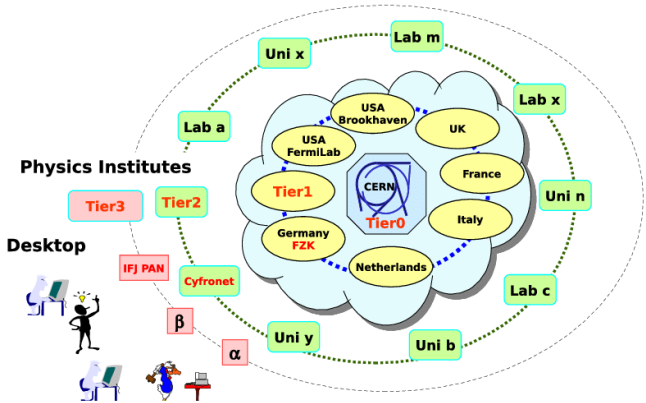


[graphics by J. Catmore]

- Core in C++, glue (job options) in Python
- Code in open gitlab repository at CERN
- User modifies job options (Python) and writes analysis algorithms (C++)
- Users typically use Athena on DAODs to produce ntuples on the grid
- Athena framework also used for all production jobs

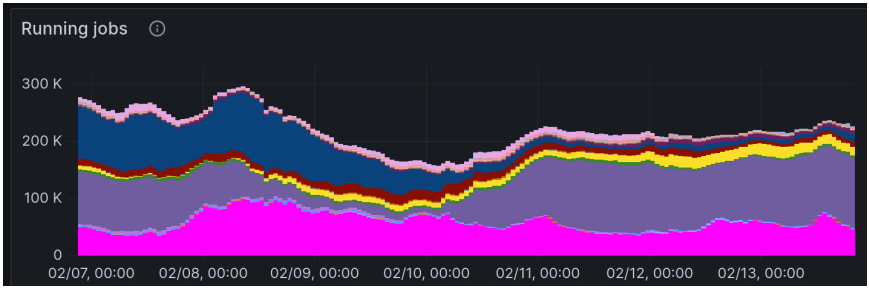
Worldwide LHC Computing Grid

- Collaboration of $\sim$ 170 computing centers worldwide
- Partnered with
  - European Grid Infrastructure (EGI)
  - Open Science Grid (OSG)
  - Nordic e-Infrastructure Collaboration (NeIC)
- Organization in Tiers 1 – 3
  - UNI-SIEGEN-HEP is Tier-3
- Jobs run where data reside, but data also move to idle sites



[Comp.Sc. 9 (2008) 47-54]

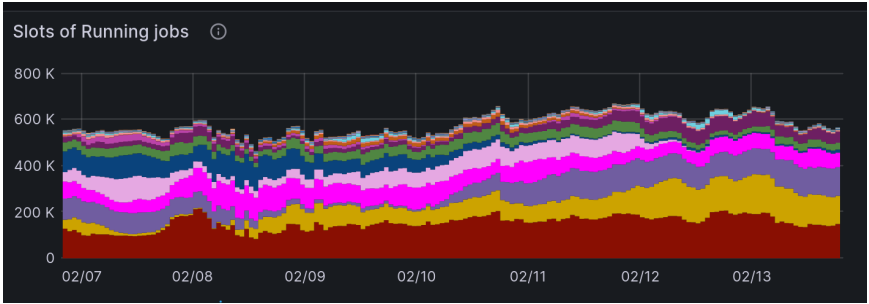
ATLAS Jobs on the Grid - by Jobs



Name	Mean	Max	Min
MC Other evgen	75.4 K	124 K	0
User Analysis	58.0 K	98.9 K	0
MC 16 evgen	39.1 K	113 K	0
MC 16 simul	11.3 K	18.6 K	0
MC Derivations	10.2 K	20.9 K	0
Data Derivations	5.79 K	13.3 K	0
MC Other	4.00 K	6.90 K	0
MC 16	3.61 K	7.32 K	0
Express	2.02 K	3.46 K	0
MC Other simul	1.89 K	5.67 K	0

- About $\sim$ 250 to 300 K jobs running concurrently past 7 days
- Mostly Monte-Carlo production and then user analysis
- Note: ATLAS has not been taking new data since November last year
- Production jobs often run in multi-core configuration, while many user analysis jobs run single-core

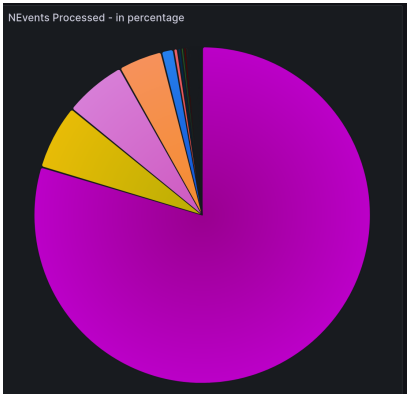
ATLAS Jobs on the Grid – by Slots



	min	max	avg
MC 16 simul	0	215 K	152 K
MC Derivations	0	173 K	82.1 K
MC Other evgen	0	124 K	75.7 K
User Analysis	0	112 K	72.3 K
Data Derivations	0	107 K	46.9 K
MC 16 evgen	0	115 K	43.9 K
MC Other	0	56.0 K	36.0 K
MC 16	0	66.7 K	31.6 K
MC Other simul	0	30.1 K	8.40 K

- About $\sim$ 550 to 600 K cores occupied during past 7 days
- Mostly Monte-Carlo simulation (typically multi-core)

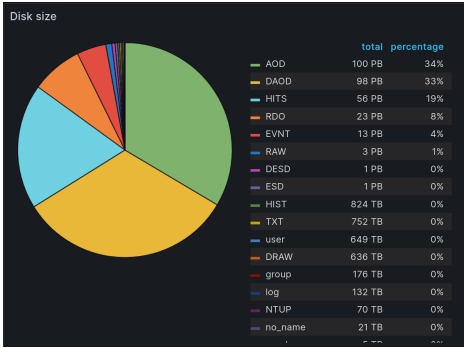
ATLAS Jobs on the Grid – Events Processed

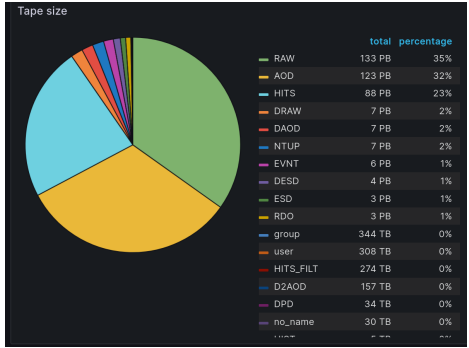


	Value	Percent
User Analysis	398 Bil	81%
MC Derivations	30.9 Bil	6%
Data Derivations	29.5 Bil	6%
Express Analysis	20.7 Bil	4%
MC merge	5.51 Bil	1%
Express	1.55 Bil	0%
MC 16 evgen	1.26 Bil	0%
MC Other	854 Mil	0%
MC 16 simul	843 Mil	0%
Group Analysis	383 Mil	0%

- Almost 400 billion events processed during past 7 days
- More than 75% by user analysis jobs
- User analysis and derivation jobs run faster

ATLAS Data on the Grid – Disk and Tape

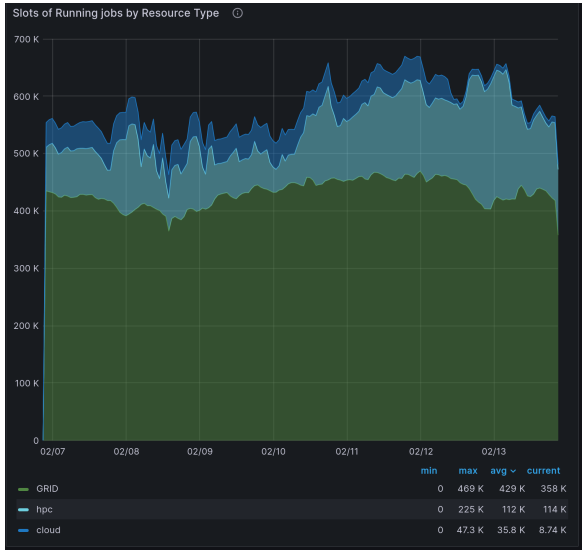




- Almost 300 PB on disk, about 380 PB on tape (total of 680 PB)
- Analysis related data (DAOD, AOD, HITS) dominate on disk
- Raw detector data (RAW) are kept on tape, but also AODs and HITS data

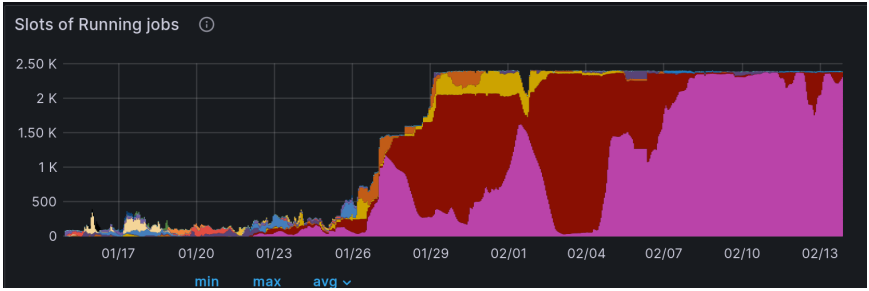
ATLAS Jobs – by Resource Type

- Slots of running jobs by resource type during past 7 days
- Besides classical Grid:
  - HPC centers ($\sim$ 110 K cores)
  - Cloud resources ($\sim$ 35 K cores)



UNI-SIEGEN-HEP WLCG Instance

Universität Siegen



	min	max	avg
MC Other evgen	0	2.37 K	800
MC 16 evgen	0	2.32 K	519
MC 23 simul	0	341	48.5
MC 16 simul	0	379	26.1
Express	0	132	16.1
MC 23	0	219	12.5
Data Derivations	0	305	8.42
MC merge	0	125	6.36
MC Other	0	69.3	6.25

- $\sim$ 100 slots at ENC, $\sim$ 2300 slots at NDC
- Plot of running jobs during last 30 days
- Running multicore jobs, mostly Monte-Carlo production
- Network bandwidth limited at NDC
  (currently 1 Gb/s connection to world, hardware for faster connection available, waiting for ZIMT)

User Analysis Resources

Preparation of analysis-specific datasets:

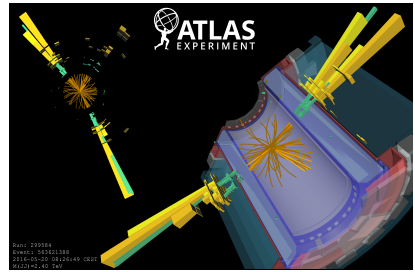
- WLCG GRID (*analysis jobs* for ntuple production)

For data analysis steps:

- lxplus at CERN
- National Analysis Facility (NAF, DESY)
- OMNI cluster (ZIMT, to run e.g. ML jobs)
- SiMPLE cluster

ATLAS environment locally:

- ATLAS software via CVMFS
- Centos7 or AlmaLinux environment via apptainer container
- Locally installed versions of ROOT
- Tensorflow and Co. typically in virtual environments



SiMPLE Cluster

SiMPLE
Siegen MultiProcessor Linux Environment

Cluster of HEP group

- Located in ENC-A 104
  - own HEP subnet (141.99.204.0/23)
- 364 cores (compute nodes) + $\sim$ 160 cores (PCs)
- $\sim$ 550 TB on storage servers (RAID-6)
- OS: Debian 12.5 (bookworm)
- Central installation system: FAI & Puppet
  - all compute nodes and PCs run exactly same software
- Central home directories
  (on redundant hardware, multiple levels of backups)
- SLURM batch system
- Experiment specific software (ATLAS, Auger, ...)

Conclusions

Detector data processing:

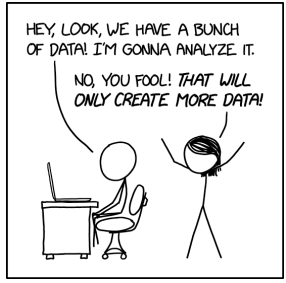
- Complex selection of interesting processes by trigger system
- Reconstruction of physics objects on the Grid

Monte-Carlo production:

- Various MC generators for physics processes
- Detector simulation is CPU & time consuming
- Same reconstruction algorithms run as for data
- All production steps performed on the Grid

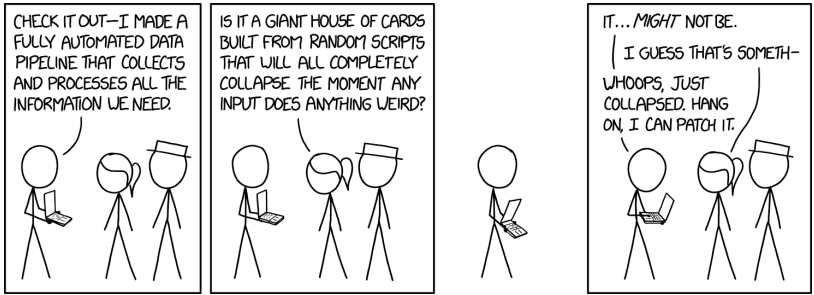
User analysis:

- Analysis specific datasets derived (data and MC)
- User typically produces ntuples on the Grid
- Runs analysis locally



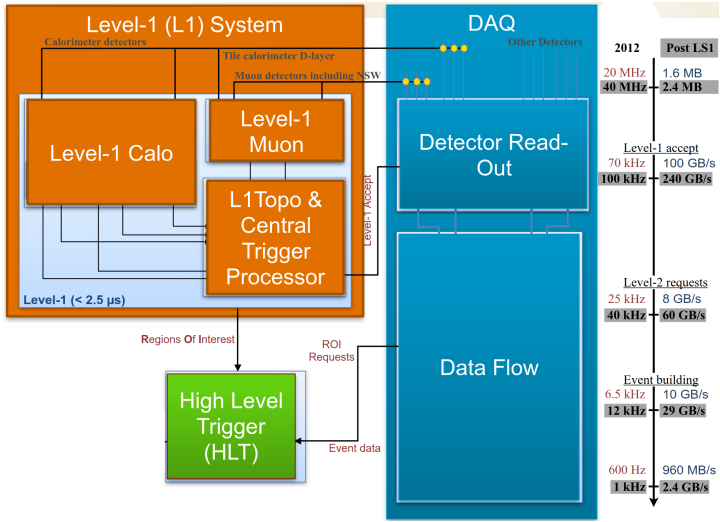
[xkcd/2582]

Bonus Material



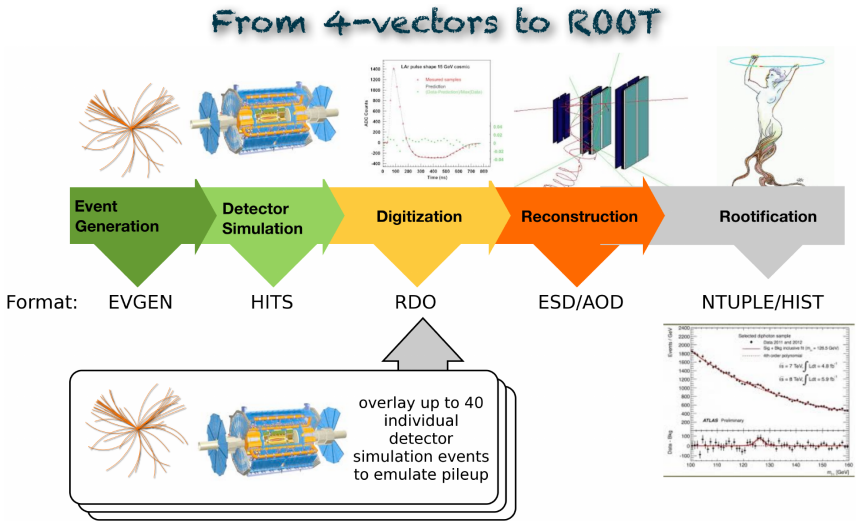
[xkcd/2054]

ATLAS TDAQ System



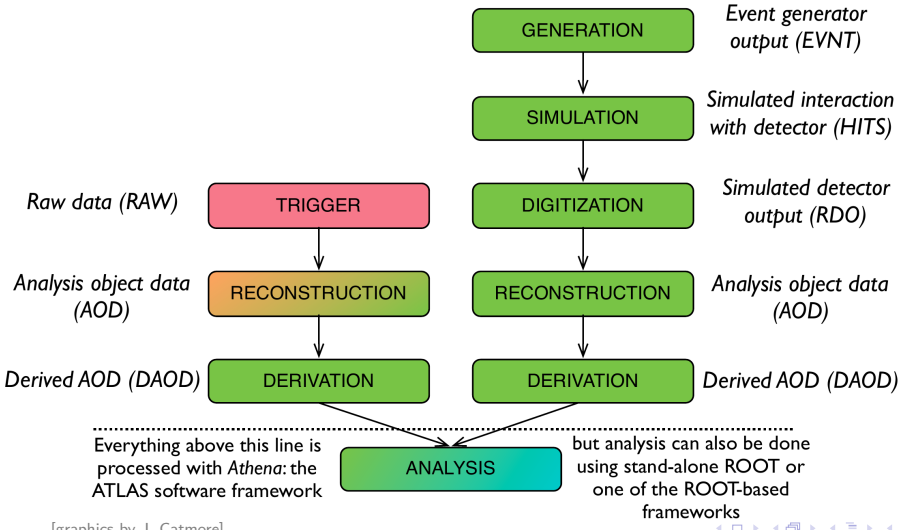
[graphics by C. Bernius]

ATLAS Monte Carlo Production Chain



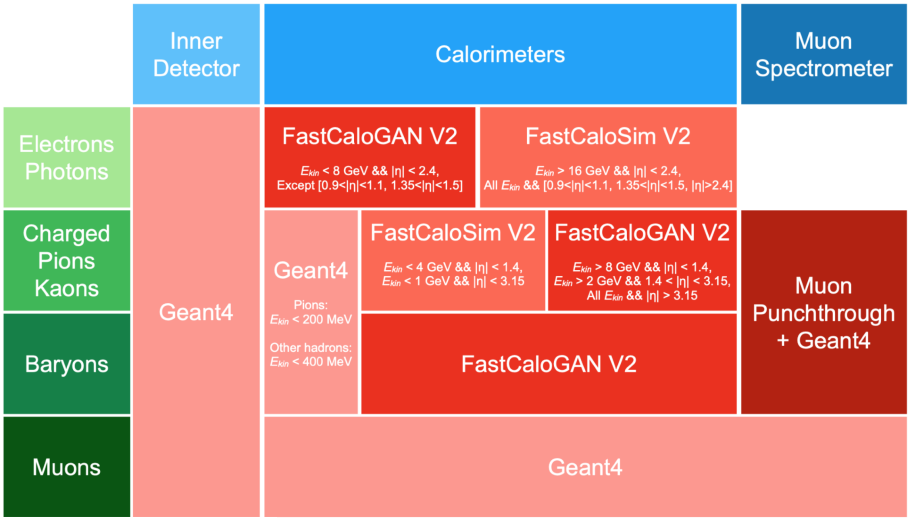
[graphics by C. Debenedetti, CHEP 2013]

ATLAS Data Formats



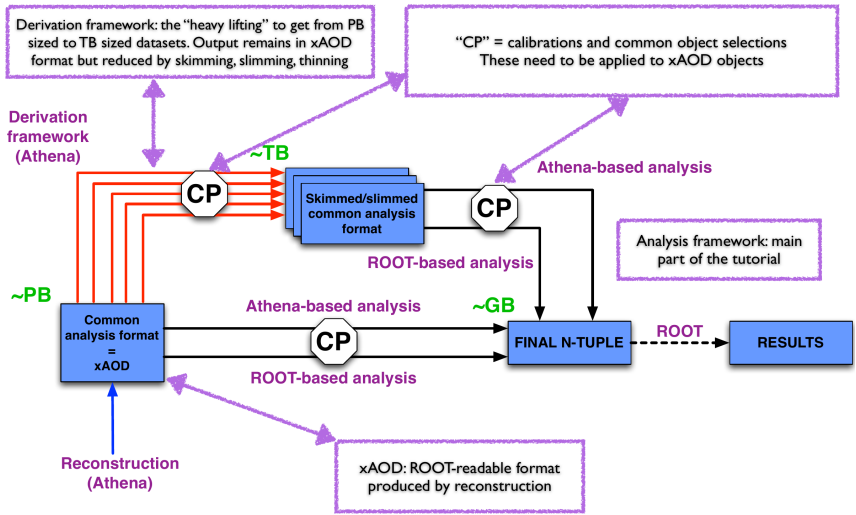
[graphics by J. Catmore]

Run 3 AtlFast3 Configuration



[J. Dandoy, ATLAS week Feb. 2024]

ATLAS Analysis Model



[graphics by J. Catmore]