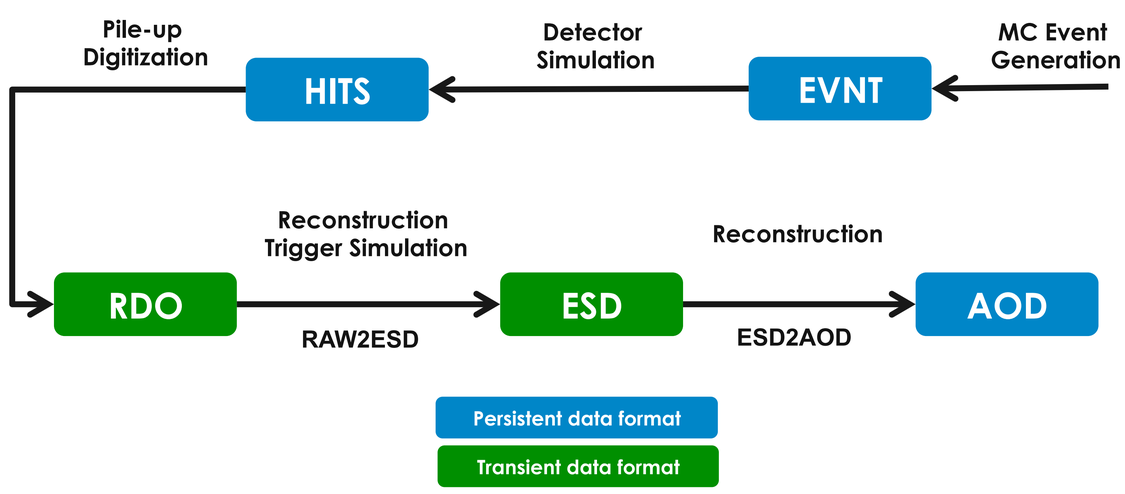
Challenges of the ATLAS Monte Carlo Production during Run-I and Beyond

J. Chapman[1], W. Ehrenfeld[2], J. Ferrando[3], J. Garcia Navarro[4], C. Gwenlan[5], S. Mehlhase[6], V. Tsulaia[7], A. Vaniachine[8] and J. Zhong[5] for the ATLAS Collaboration

[1]University of Cambridge, [2]Universitaet Bonn, [3]University of Glasgow, [4]Instituto de Fisica Corpuscular, [5]University of Oxford, [6]Ludwig-Maximilians-Univ. Muenchen, [7]Lawrence Berkeley National Lab. , [8]Argonne National Laboratory

- **Event Generation :**
  - Around 30 Monte-Carlo generators used to simulate physics processes :
    - Framework integrated generators and stand-alone generators
  - Event generation workflows :
    - Single step generation: Pythia6/8, Herwig(++), Sherpa
    - Two-step generation: parton level generator (Alpgen, MadGraph, …) coupled via LHEF files to framework generator for hadronisation
- **Simulation :**
  - Geant4 (G4) full simulation:
    - all stable particles are tracked through the ATLAS geometry
  - Geant4 full simulation with Frozen Showers in calorimeters: 25% speed up
    - showers are tracked down to very low energy by Geant4 → stop showering at a threshold and substitute by a pre-made list of energy deposits
  - AtlFast-II (AF-II): factor 10 speed up in mc12
    - parameterize all particles except muons in the calorimeters
  - Integrated Simulation Framework (ISF)
    - better integration of full and fast simulation based on sub-detectors and particles

ATLAS Monte Carlo Simulation Flow

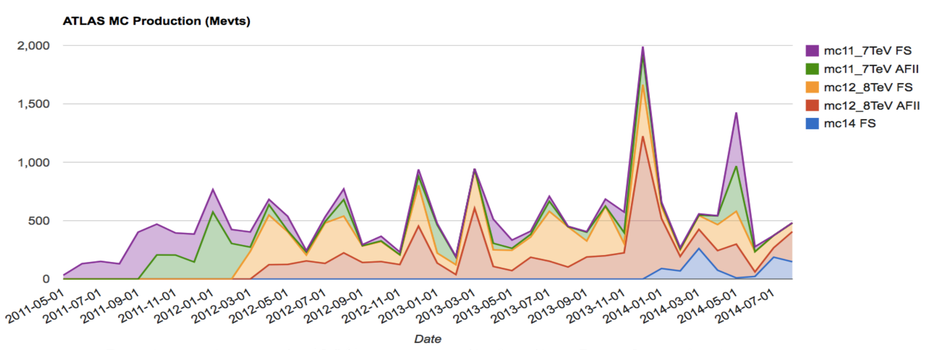


- **Digitization :**
  - Simulate detector readout
  - Simulate pile-up contributions (multiple pp interactions on top of hard scatter event)
  - Overlay a number of pre-simulated minimum bias events on each signal event
  - Optimize pile-up event storage and access
    - Cache pile-up events in memory → memory intensive
    - Flush memory early and re-load from disk on demand → I/O and CPU intensive
- **Reconstruction :**
  - Reconstruct simulated events in the same way as data
  - Trigger simulation
  - Two step process:
    - RAWtoESD: main reconstruction → output is Event Summary Data (ESD)
    - ESDtoAOD: fast slimming process → output is Analysis Object Data (AOD)
  - D3PDs Ntuples and derived formats from ESD or AOD are produced by Group Production

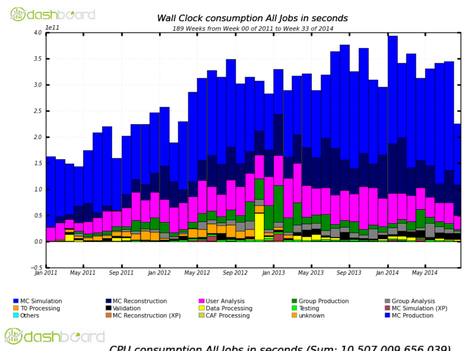
MC Campaigns

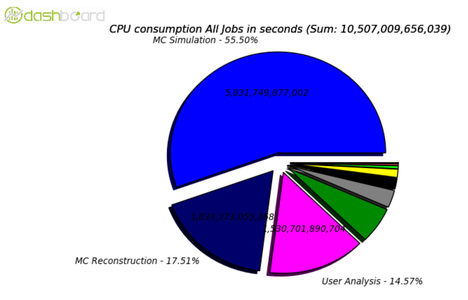
- MC production campaigns correspond to data taking periods with same conditions
  - center-of-mass energy, detector configuration, conditions, ...
- Major MC production campaigns in Run-1 :
  - **mc11**: simulation configuration for 7 TeV in 2011. Four different sub-campaigns. implementing : pile-up condition, detector conditions and geometry closer to data.
  - **mc12**: simulation configuration for 8 TeV in 2012 . Three sub-campaigns. Main part of the events were produced in sub-campaign mc12b. Late mc12c implemented an improved detector geometry description.
  - **mc14** : preparation for Run-2 production:
    - 8 TeV : Improved and updated simulation, digitization and reconstruction with same conditions as mc12 campaign
      - ISF framework used as main simulation framework.
    - 13 TeV : Campaign with the center of mass energy expected for Run-2 and estimate of the pile-up and detector conditions.
      - Multicore processing becomes default for production (simulation, digitization and reconstruction)

Campaign	Full Simulation (10^6 events)	Fast Simulation (10^6 of events)
mc11	3640	3270
mc12	6370	6430
mc14	850	-

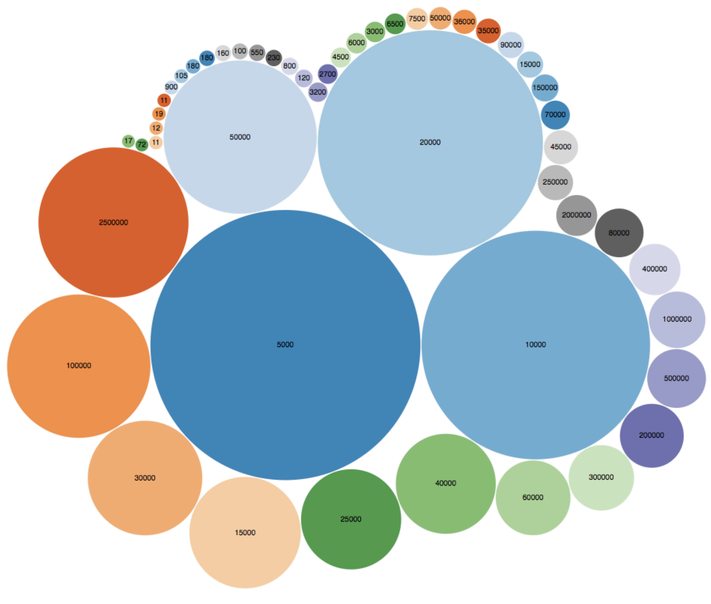


Produced events (10^6) per month for the Run-1 period





CPU consumption for the different steps of the production of simulated events



Sizes of the more than 22.000 different datasets produced during Run-1 (mc12 campaign)

Production Characteristics (mc12 example)

- **Simulation** :
  - Full simulation: 100 events per job → ~80 MB output file size → merged up to 1000 events (0.8 GB file size) for better grid transfers and tape storage
  - Fast simulation: 1000 events per job → ~0.5 GB output file size
  - low memory requirement: ~1 GB
  - run time per event (averaged over grid CPUs) :
    - G4 full simulation (335 s), G4 full simulation with frozen showers (250 s) and AtlFast-II (20 s)
- **Digitization and reconstruction :**
  - Processing 500 events per job → ~220 MB output file size → merged up to 5000 events (~2.2 GB file size) for better grid transfer and tape storage
  - High memory usage: 3.6 – 3.8 GB (in 32 bit)