Most important workloads and how to run them

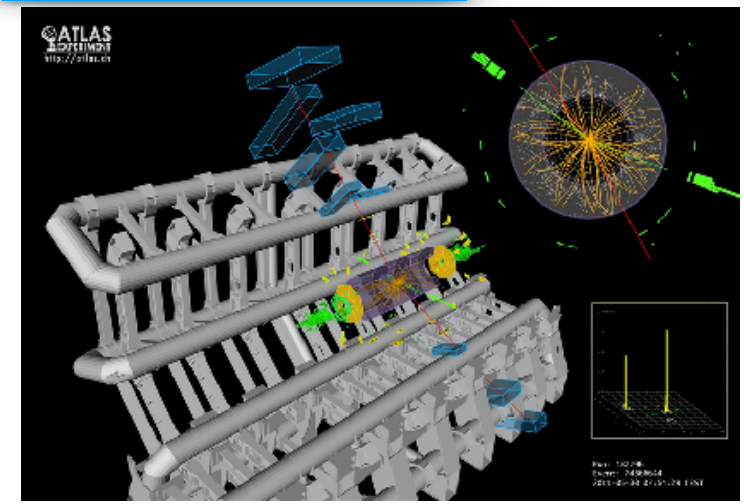
Johannes Elmsheuser

28 March 2018

Joint WLCG&HSF Workshop 2018, Naples

Introduction

1 pp-collision event:



1 event:

Array of objects with sub-detector infos

- Calorimeter
- Inner detector
- Muon detector
- ...

Array of objects with kinematic infos of physics objects

- Electrons
- Muons
- Jets
- ...

Collision events are independent

1 ROOT file:

Array of events:

...

Simulation

EVNT — *Generation*

HITS — *Simulation*

RDO — *Reconstruction*

Data

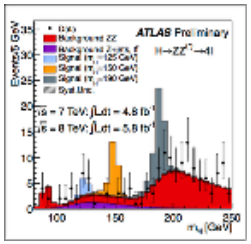
RAW

AOD — *Derivation/Filtering*

xAOD — *Analysis*

used in statistical analysis of many events

ROOT file formats:



Cost and workflow examples

- Costs of workflows: **CPU, memory, disk** and **network I/O**
- Typical parameters to describe job requirements in WFMS or batch system
- We have collected examples from ALICE, ATLAS, CMS, LHCb
- Examples loosely categorised by:
  - **CPU intensive:** Geant MC simulation
  - **CPU + I/O:** MC digitisation + reconstruction
  - **I/O heavy:** event filtering/streaming, analysis
- **Memory:** Fit into $\approx$ 2 GB/CPU core grid slot (can vary)
- **Network:** not directly specified in examples - input files usually locally read, only remote conditions DB access

Notes on the workflows I

MC simulation:

- Event sizes: input ≈O(1-3 MB), output ≈O(1 MB)
- Processing times: O(30s-several min) (Geant)
- Large fraction of all experiment grid walltime goes into Geant simulation, e.g. ≈70% for ALICE or ≈ 50% ATLAS
- CPU/Walltime efficiency for simulation usually rather good 80-95%,

Reconstruction:

- Event sizes: input ≈O(1 MB), output ≈O(0.1-0.5 MB)
- Processing times: O(10-30s)
- For higher average number of interactions per bunch crossing ($\rightarrow$Run4), reconstruction will take much longer and therefore requires higher share (see backup slides)

Notes on the workflows II

Event filtering/Analysis:

- for I/O intensive jobs CPU/walltime eff can drop significantly also depending on local disk or streaming from SE

How do things scale with pileup (example from CMS)?

- event generation and simulation independent of pile-up
- Digitization: pileup input size scales ≈linearly. CPU time and output size scale less than linearly
- Reconstruction: CPU time scales worse than linearly. Output sizes scale linearly or less than linearly.

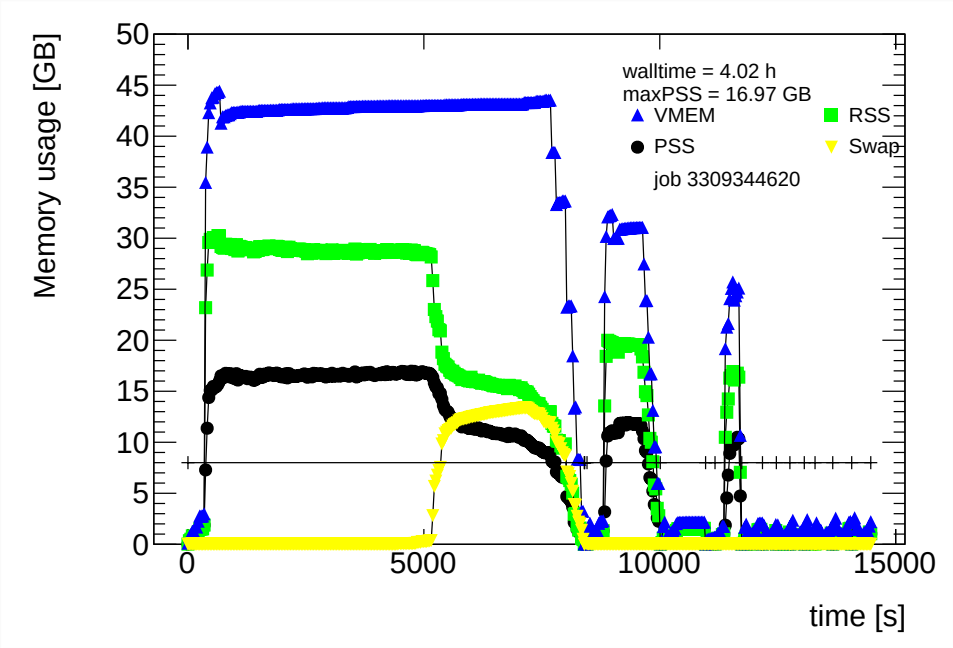
Example workflows

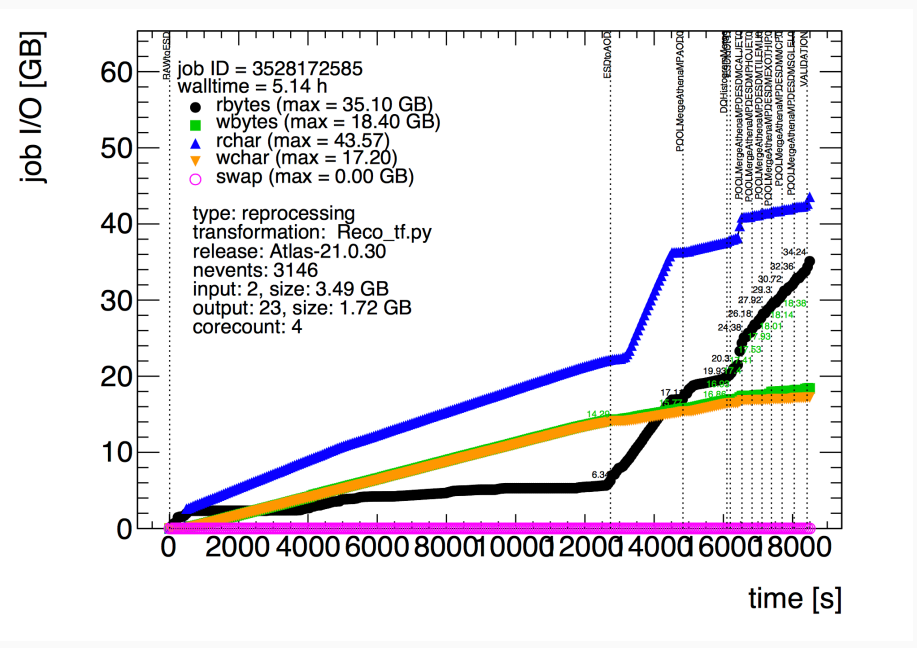
- Instructions available how to run (link to google doc):
  - SLC6+CVMFS+input file + (voms proxy) + executable wrapper in python/bash with configuration
  - Docker containers (also used in WLCG benchmarking WG) for all experiments
- LHCb:
  - Docker container: lhcb-montecarlo-demo
- ALICE:
  - Docker container or direct example at ALICE pp simulation example code
- CMS:
  - Full MC chain: event generation, simulation, digitization, reconstruction, analysis data creation - see next page
- ATLAS:
  - 3 typical examples: Geant4 simulation, MC digitization+reconstruction, DxAOD derivation production

CMS EVENT GENERATION AND SIMULATION

```
#set up the software area
source /cvmfs/cms.cern.ch/cmsset_default.csh # or sh

#set up a recent release
scram p CMSSW_10_0_1
cd CMSSW_10_0_1
cmsenv

#a cms grid certificate needed when not running at CERN
voms-proxy-init -voms cms

#run the generation and simulation (TTbar events at 13 TeV and 2017 CMS detector)cmsDriver.py)
 cmsDriver.py TTbar_13TeV_TuneCUETP8M1_cfi  --conditions auto:phase1_2017_realistic -n 100 \
  --era Run2_2017 --eventcontent RAWSIM --relval 9000,50 -s GEN,SIM --datatier GEN-SIM \
  --beamspot Realistic50ns13TeVCollision --fileout step1.root --nThreads 8
```

Profile of single ATLAS jobs





- Memory: PSS, RSS, VMEM, Swap
- Disk I/O: rbytes, wbytes, rchar, wchar

Conclusions

- ALICE, ATLAS, CMS, LHCb provided examples how to run workflows for CPU and I/O intensive MC data processing
- Costs of workflows: **CPU, memory, disk** and **network I/O** which are the typical parameters to describe job requirements in WFMS or batch systems

BACKUP

Reconstruction walltime per event vs. <μ>