• Corentin Bugnot’s 2018 studies on reference HEP application’s behavior under emulated resource conditions
  • Focus on overall performance (runtime)
• PRmon detailed behavior of core metrics over time

• Now: Combining the two approaches to get input for modelling the behavior of individual metrics under resource restrictions
• In addition parametrization of PRmon timeseries as series of linear approximations

**TEST SETUP**

- 64 GB DDR4 Memory
- Dual Socket, 8 Core, 2HT
- Total: 32 logical cpus
- Intel PCIe SSDs for local disk IO

**Reference workloads:**

- **ATLAS:** Single job, 16 Processes
  1. Monte Carlo Generation: Memory
  2. DigiReco: Memory, Bandwidth, Latency
  3. Deriv + Prod: Memory, Bandwidth, Latency

- **CMS:** Single job, 16 Threads
  1. Monte Carlo Generation: Latency
  2. Digi + Trigg + PileUp Sim: Memory, Latency
  3. Reco + Analysis: Memory, Latency

- **Resource restriction range**
  - Network bandwidth (1 to 1250 MBps)
  - Network latency (0.1 to 64 ms)
  - Memory (4 to 64 GB)
OUTPUT DESCRIPTION

• Each plot represents a job restricted in one of the mentioned resources
• Each color represents a different resource limit
• Different resource measurements:
  • Memory
    • PSS
    • RSS
    • Swap
    • Vmem
  • Network
    • rx_bytes
    • tx_bytes
    • rx_packets
    • tx_packets
  • Application
    • rchar
    • wchar
  • Disk
    • read_bytes
    • write_bytes
ATLAS WORKLOADS
**ATLAS G4 MC: Memory**

- Only after decreasing memory to 4 GB, swap is utilized
  - Job insensitive to memory limitation
- Disk and network usage increase slightly when restricting memory
  - 4GB limit causes 1GB more disk reads
ATLAS DIGIRECO: Memory

- Swap activity starts from 16GB
- In the case of the 4GB restriction, the application reads 600GB of data from disk.
- There are very high differences with respect to disk reading between the various restrictions, while the amount of data requested from the application is roughly the same.
ATLAS DerivProd: MEMORY

- Swapping starts from the 16GB restriction
- 4GB and 8GB have very similar swap behaviors, yet very different disk usage, with 4GB reading 100GB more from disk than the 8GB restriction
ATLAS DigiReco: Bandwidth

- Plots show shift in time series as we increase bandwidth limit

- The workload reads from disk more than what the application asks for. e.g. for 4 and 8 GB restrictions the application requests for ~12.5 GB but the job reads ~17 GB.
**ATLAS**

**DerivProd: Bandwidth**

- The slopes of the curves are similar even with varying limits, while we expect lower gradients with less bandwidth.
- The only effect is that IO activity seems delayed.
- The time shift causes the worst case scenario to take about twice as long than the fastest run.
ATLAS DigiReco: Latency

- The plots show a time shift as well, but slopes change.
- Time dilation is more evident: the best case scenario takes ~1800 s, the worst case takes ~3700s
ATLAS DerivProd: Latency

- There are noticeable changes in the slopes of the curves.
- The worst case scenario takes more than 10 times to complete with respect to the best case scenario.
CMS WORKLOADS

Riccardo Maganza
• 4GB and 8GB restrictions show very similar swap patterns

• However the 4GB restriction reads 80GB of data from disk
CMS RecoAnalysis: Memory

- Swaps only starts at 4GB
- There is only a ~50 second (~15%) difference between the lowest limitation and the highest one.
CMS G4 MC: Latency

- The workload shows a ~30% time increase between best and worst case scenario.
CMS DigiTriggPile UpSim: Latency

- Curves show very slight slope changes, and mostly time shifted behaviours
CMS
RecoAnalysis: Latency

- The curves again only show time shifting of similar patterns. In this case there is a 2x difference between worst and best case scenarios.
NEXT STEP

- Apply parametrization to all time series
- Include impact of restrictions
PRELIMINARY RESULT

Riccardo Maganza
BACKUP SLIDES
QUESTIONS?

riccardo.maganza@cern.ch
@RiccardoMaganza