ATLAS SCALE-UP TEST ON PIZ DAINT

Gianfranco Sciacca

AEC - Laboratory for High Energy Physics, University of Bern, Switzerland

LHConCray WG - 7 November 2017

PREPARATION

- **Panda queue created: CSCS-LCG2-HPC_MCORE_TEST**
  - **pointing to arc04.lcg.cscs.ch, queue atltest, corecount=16**
  - **reservation with 11 nodes set up by Miguel**
- **Sent some HC jobs, helped Miguel tuning the ARC conf**
- **Simulation task created by ATLAS: https://bigpanda.cern.ch/task/12491843/**
  - **4M events, 40k input files, up to 148MB/file (mostly 115MB)**
  - **jobs tuned to ~1h duration (maxEvents=100)**
  - **ramCount=900 MBPerCore**
  - **Output expected: ~70MB/job**
- **Started submitting jobs, 2 Nov at 4PM**
  - **jobs accounted to be using up to 32GB of mem and got killed**
  - **load spike on GPFS**
  - **removed memory limits, jobs started running**

PREPARATION

- **Identified an issue with the ARC infosys:**
  - **jobs of the wlcg partition were published correctly, jobs of the atltest partition were not.**
  - **This would break submission from the aCT, causing the system to drain every few hours**
- **After many attempts to fix it, it was decided late on Friday to switch to arc05 and have only the arcds1 do the staging.**
- **New CE host hardcoded in aCT so we did not need to wait for the ATLAS infosys to propagate the change**
  - **jobs started to run, and ran stable over the weekend, filling the 11-node allocation**
- **On account of the low memory usage, ATLAS proposed on Sunday to try out 18-core jobs in order to fill the nodes**
  - **Miguel switched to allow 72-cores per node un Sunday evening**
  - **ATLAS overrode the 16-core setting for the task directly on the aCT**
  - **18-core jobs ran stable overnight**

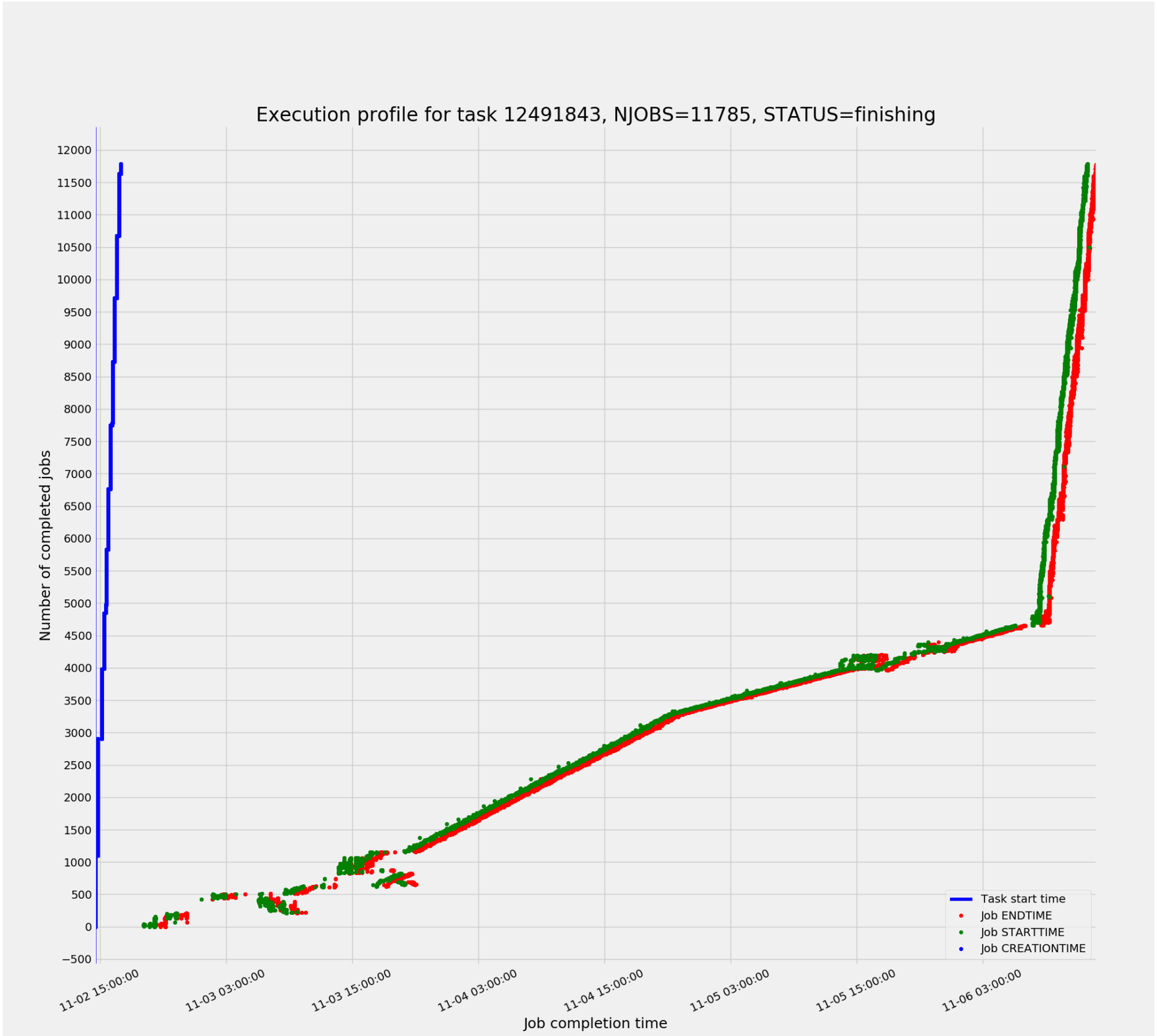
SCALING UP

Started 06 Nov 8 AM

- **Decided initially to ramp up in 5 stages in order to avoid nasties**
- **Eventually went for all-at-once**

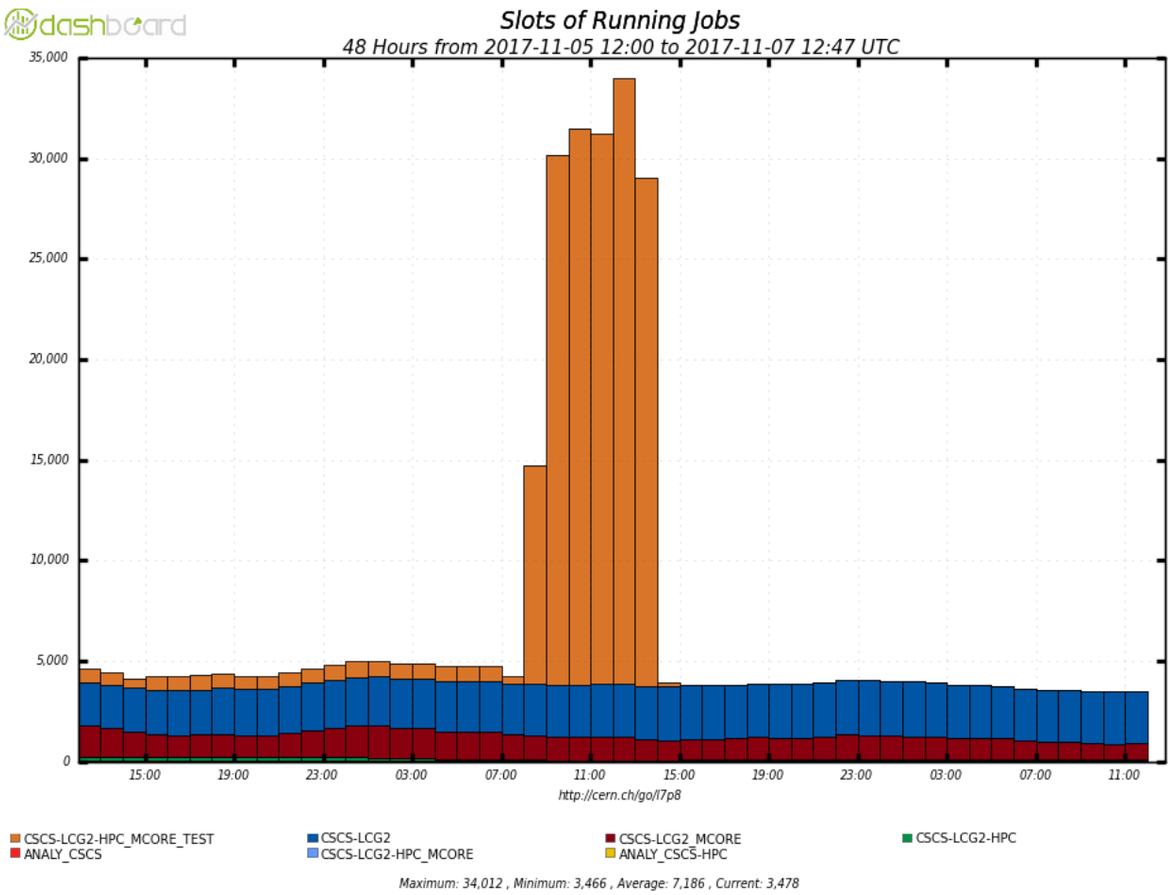
- **Reached 1420 jobs (25560 cores) in ~1h**
  - **aside from a glitch with ARC that added about ~20 min delay**
  - **fairly linear otherwise, 27 jobs/min**
  - **seemingly dominated by slurm**
- **ARC unstable, a-rex getting stuck repeatedly, needs to be restarted by hand**
  - **Realised we don't have the latest bugfix version**
  - **Upgrade on the fly vs babysit**
  - **Went for the latter, many restarts needed**
  - **Incresed the maxqueued on the aCT to have a large enough buffer and avoid draining between restarts**
- **Stable running for 3h from 11 AM**
- **Stopped submission at 2 PM**
- **Killed all running from the aCT at 14:45**
- **System clean at 3 PM**

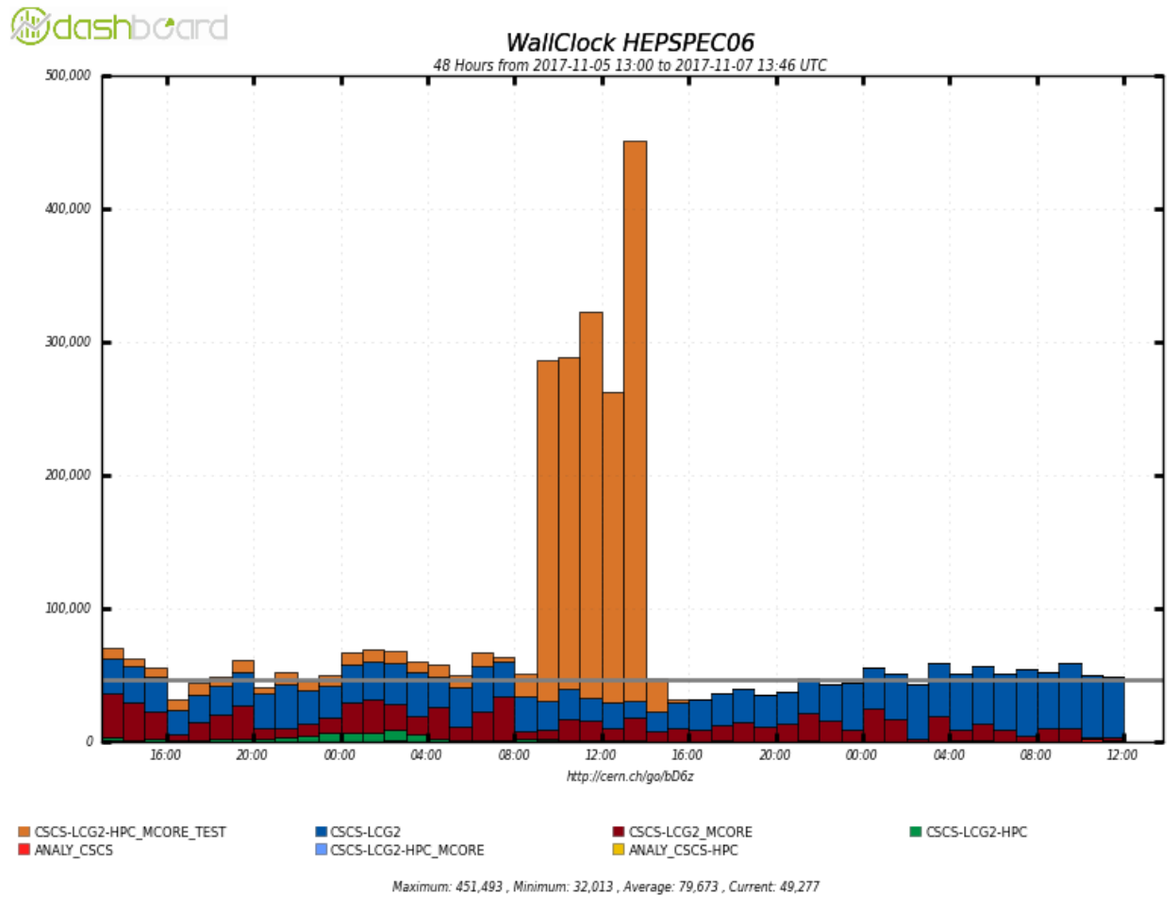
TEST SUMMARY



TEST SUMMARY

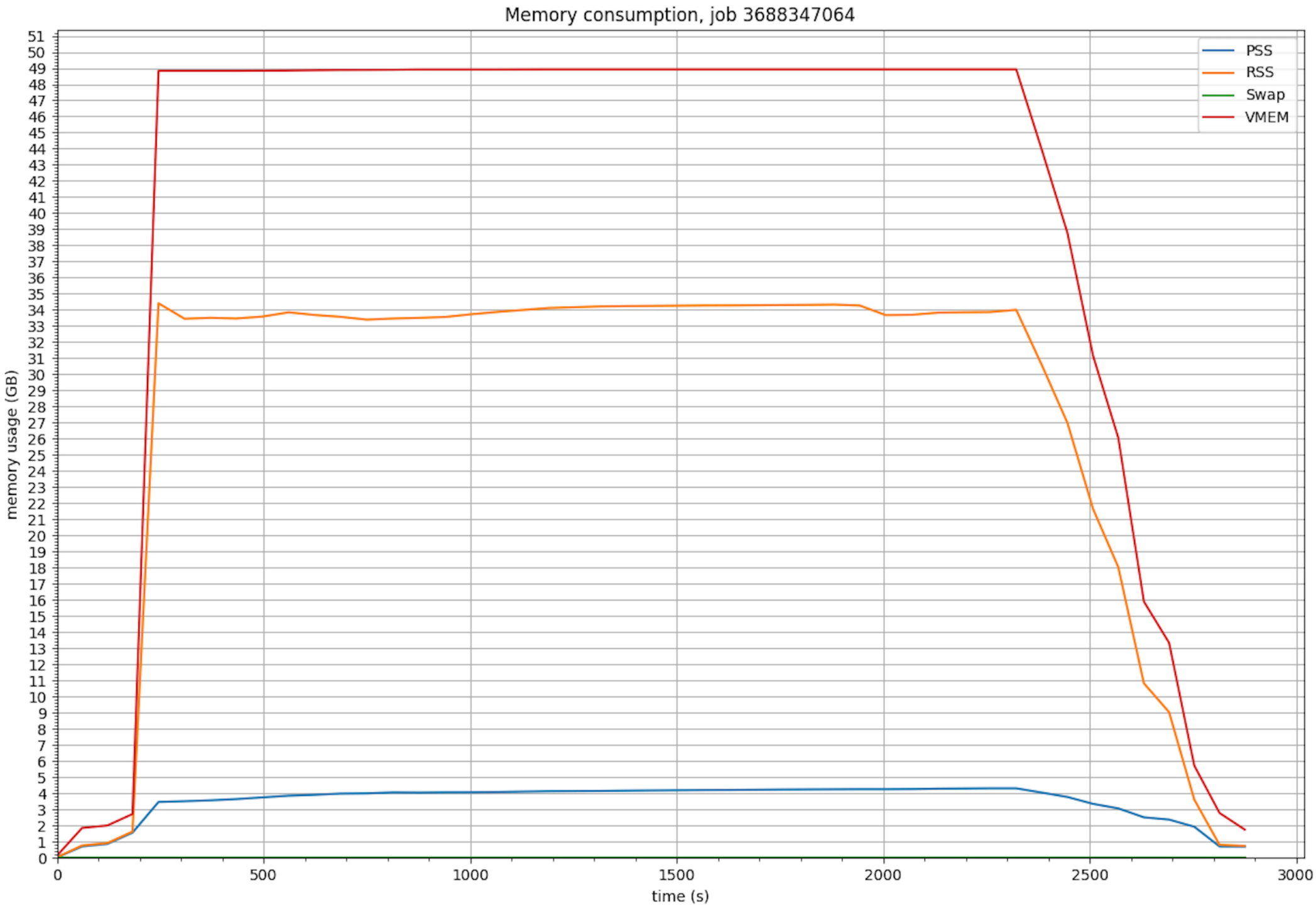
- 1M events processed (25% of total): 10162 jobs (out of 11785)
- Total input size: 1TB (no ARC caching), output size: 0.7TB (to the Nucleus in Spain)
- Max running jobs reached 1432 (25774/27648 cores - 93.22%)
- Failure rate ~1% (but all retried), CPU/WC eff 0.76 (due to artificial job length)





JOB PROFILE

https://bigpanda.cern.ch/memoryplot/?pandaid=3688347064



JOB PROFILE

https://bigpanda.cern.ch/memoryplot/?pandaid=3688347064