Opportunistic HPC usage & Whole-node scheduling

ATLAS DE Cloud F2F Meeting

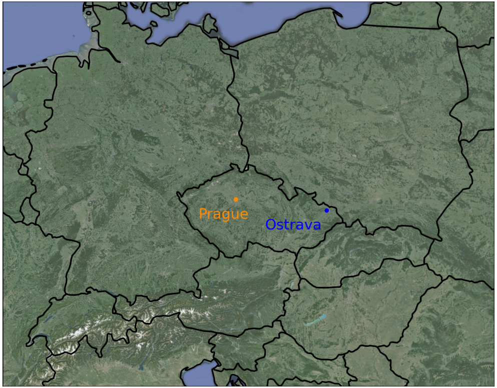
Michal Svatoš

Institute of Physics, AS CR

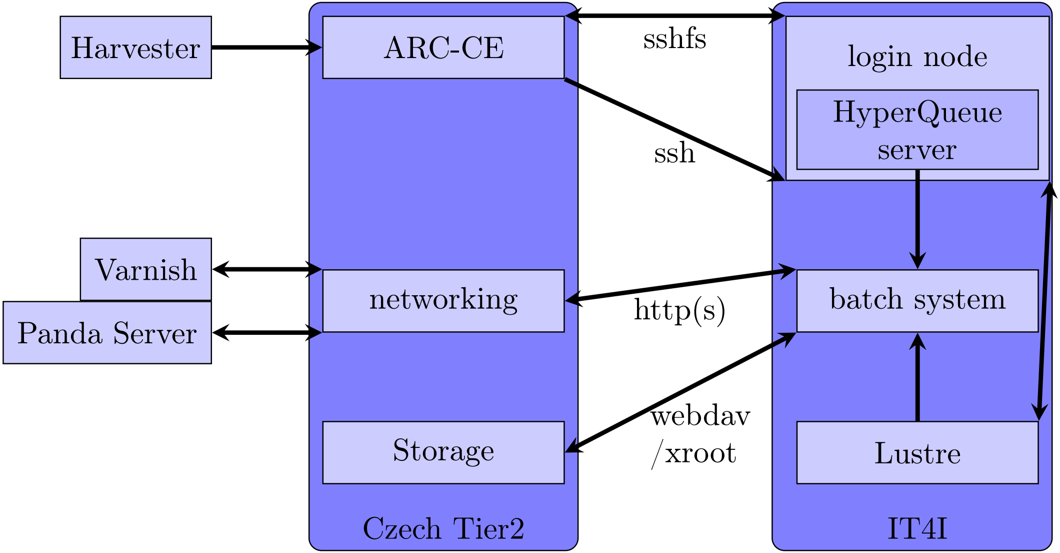
6.-7.10.2025

Introduction

- ATLAS is using resources of IT4Innovation (located in Ostrava) since 2017
- usage via praguelcg2 (278km from Ostrava)
- currently used HPCs
  - Barbora (CPU nodes: 192 WN with 36 cores and 192GB of RAM) since 2020
  - Karolina (CPU nodes: 720 WN with 128 cores and 256GB of RAM) since 2021
- both machines allow only whole-node scheduling
- both queues work in pull mode

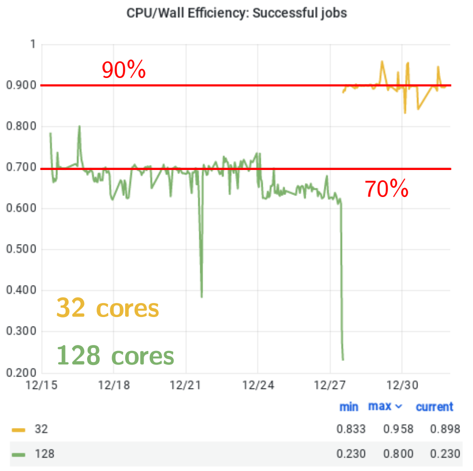


Job submission system



HyperQueue

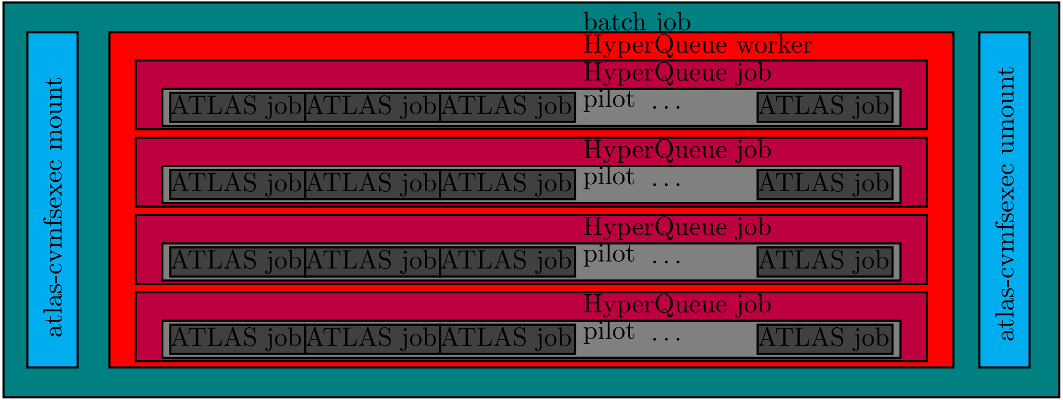
Motivation:
Karolina: the switch from one whole node (128-core) job to four 32-cores jobs was because of CPU efficiency increase:



	min	max ⌄	current
🟨 32	0.833	0.958	0.898
🟩 128	0.230	0.800	0.230

HyperQueue

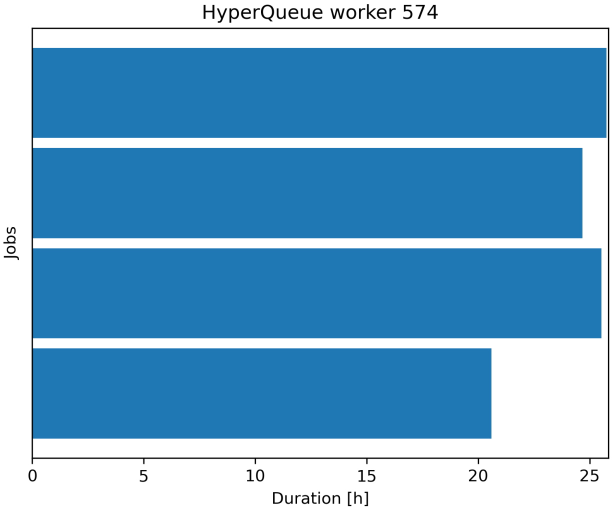
inside of Karolina batch job:



HyperQueue

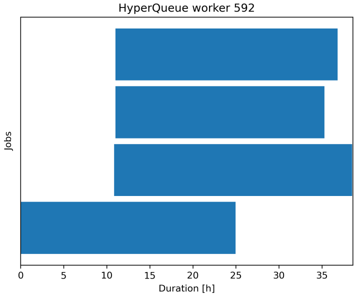
filling efficiency on Karolina:

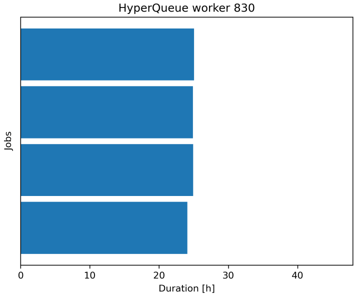
common case:



HyperQueue

filling efficiency on Karolina:
exceptional cases:





possible cause: late start could be caused by lack of jobs

- there will be a parameter not allowing batch job to be submitted until there is enough ATLAS jobs to completely fill it

possible cause: jobs killed could cause early end (if something remained stuck and keep running until max time)

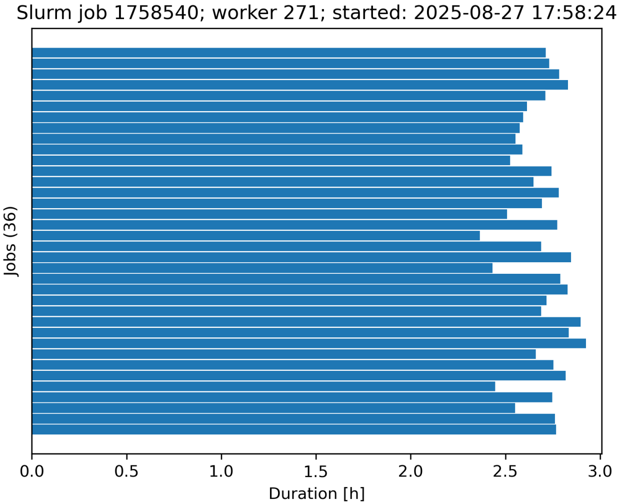
- there is an easy workaround: script, run by cron every 10 minutes, correlating running HQ jobs and HQ workers and closing empty HQ workers

Job mixture

- due to lack of sim and abundance of evgen (single core only) at the time, I started investigating job mixture at HPC
- next to `praguelcg2_Barbora_MCORE` for sim, I added `praguelcg2_Barbora_SCORE` for evgen
  - so I can have separate timefloors
  - they share ARC-CEs
- a plan is to run 36 score jobs on Barbora and up to 64 score jobs on Karolina

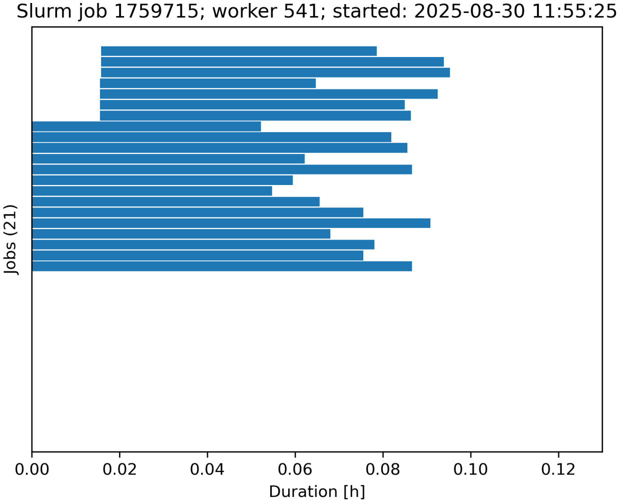
Job mixture

- some tuning and changes in HyperQueue scheduler needed
  - sometimes it works nicely



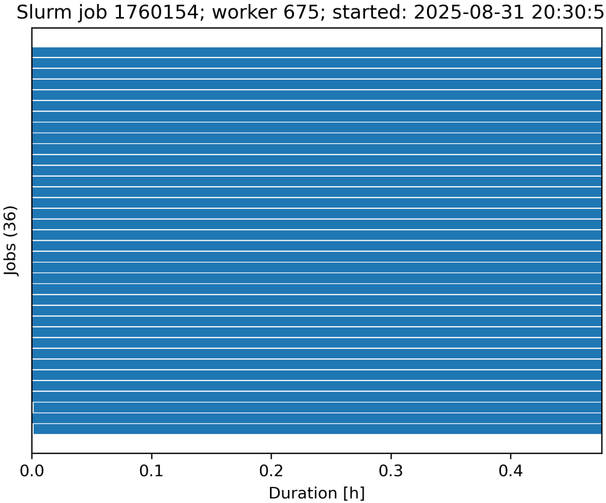
Job mixture

- some tuning and changes in HyperQueue scheduler needed
  - sometimes, not so much



Job mixture

- some tuning and changes in HyperQueue scheduler needed
  - also, we are running on pre-emptable queue (which complicates things)


Slurm job 1760154; worker 675; started: 2025-08-31 20:30:51
Jobs (36)
Duration [h]