LHCb experience with HPC centres

Alexandre Boyer (CERN), Concezio Bozzi (INFN)

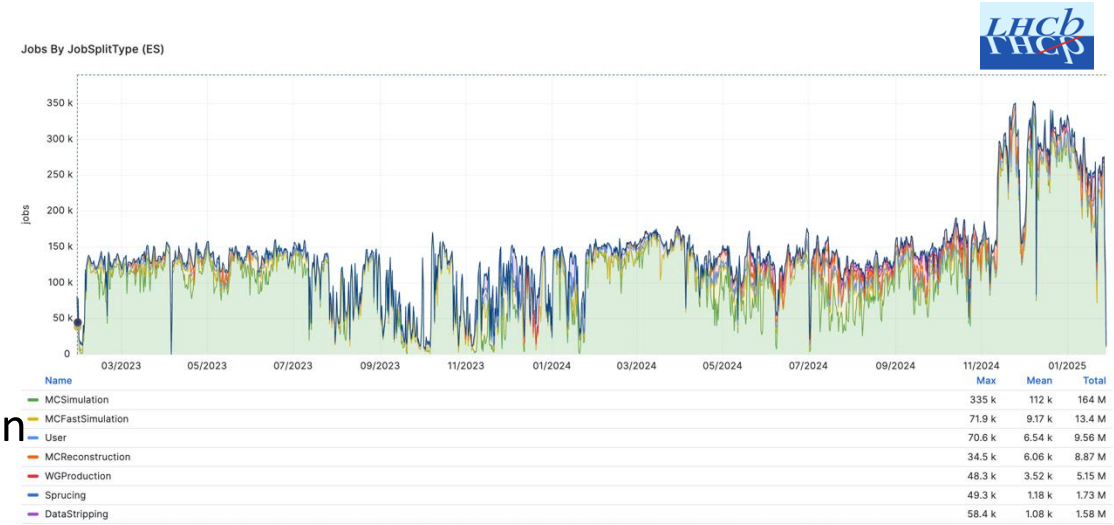
On behalf of the LHCb distributed computing team

HEP/HPC workshop, CERN, January 30th 2025

Offline activities dominated by Monte-Carlo production

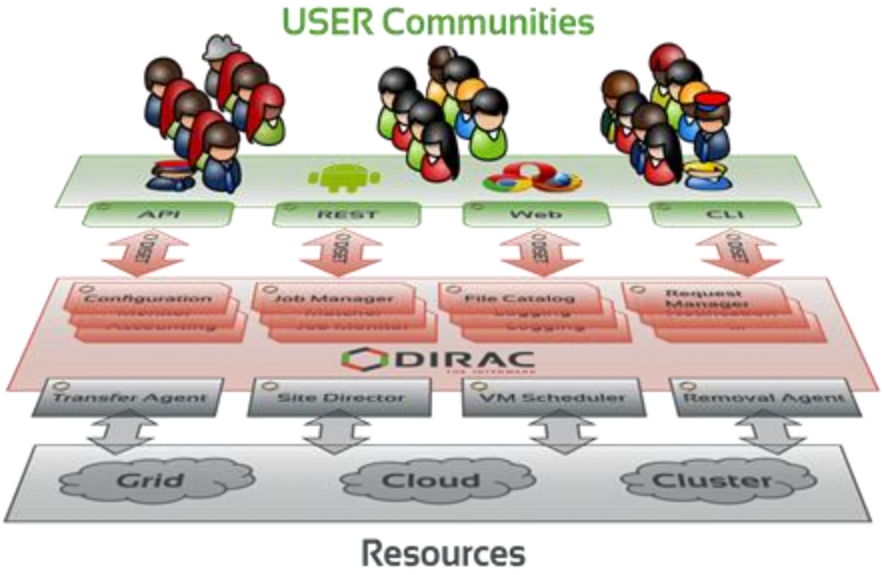
Geant4-based simulation (GAUSS)

- (almost) no input data
- CPU-intensive task
- Single-core, 2GB RAM
- Multi-threaded version in progress
- ARM support under validation
- Fast simulation options also available
  - Used to produce ~2/3 of simulated events



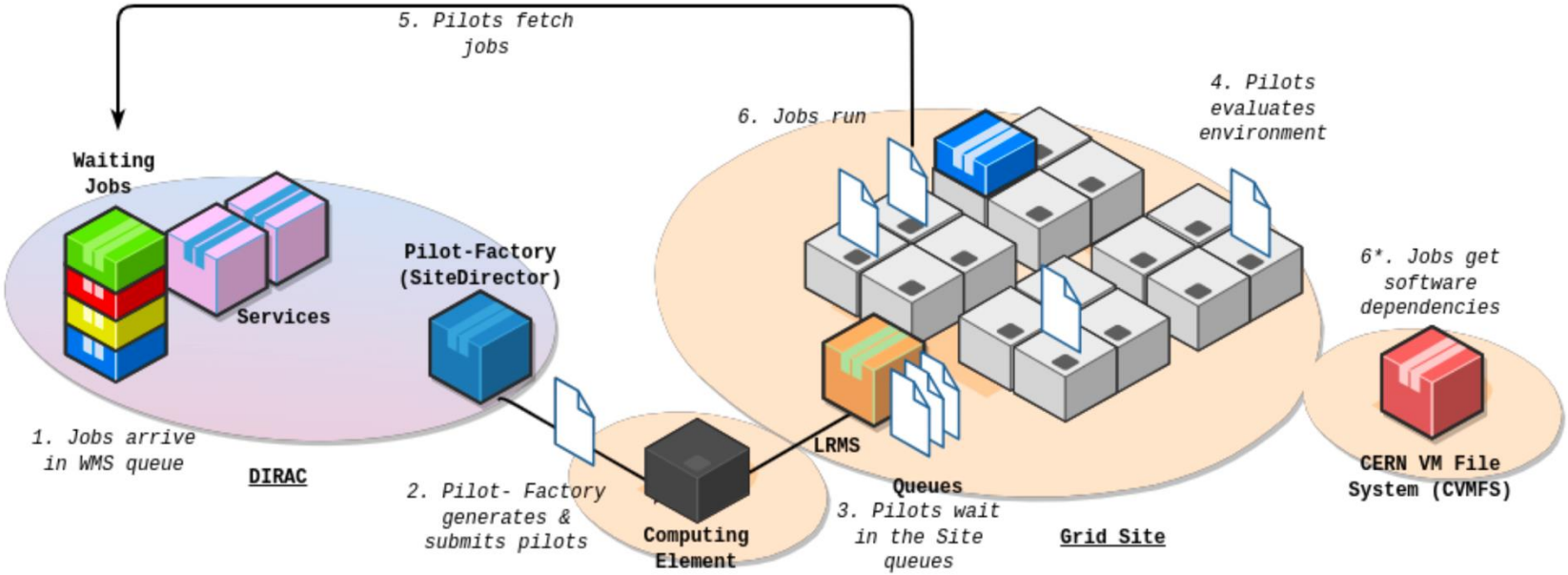
Everything through DIRAC

- Open-source middleware for distributed computing
- Workload Management System to submit jobs to remote, shared and heterogeneous computing resources
- Data Management System handling data transfer to / interact with storage interfaces
- Started as an LHCb project, it is experiment-agnostic since 2009
- Shared by multiple projects in HEP, astronomy, life science



https://dirac.readthedocs.io/en/latest/

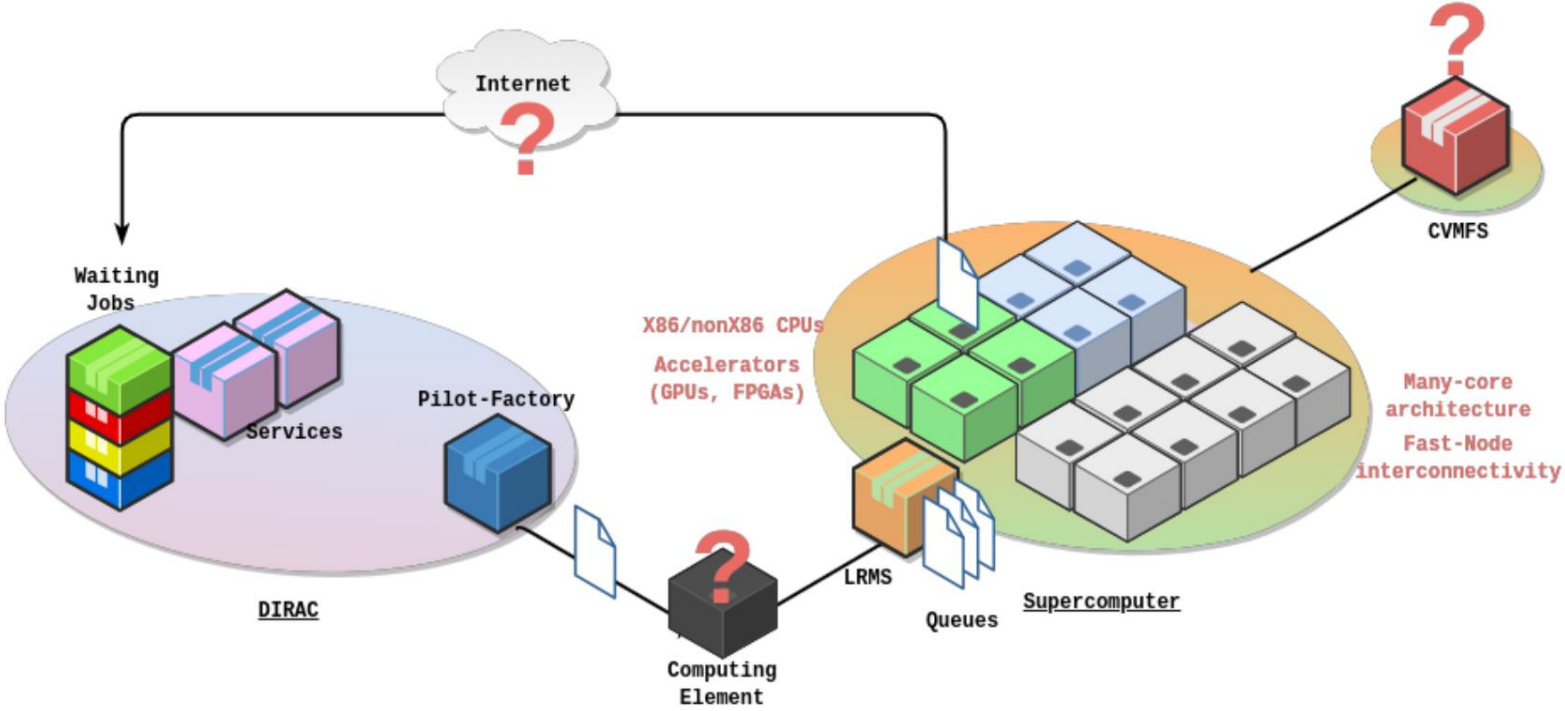
DIRAC Workload Management Service & WLCG



…getting allocations using the pull model

(c.f. CHEP2023 presentation)

DIRAC Workload Management Service & HPCs?

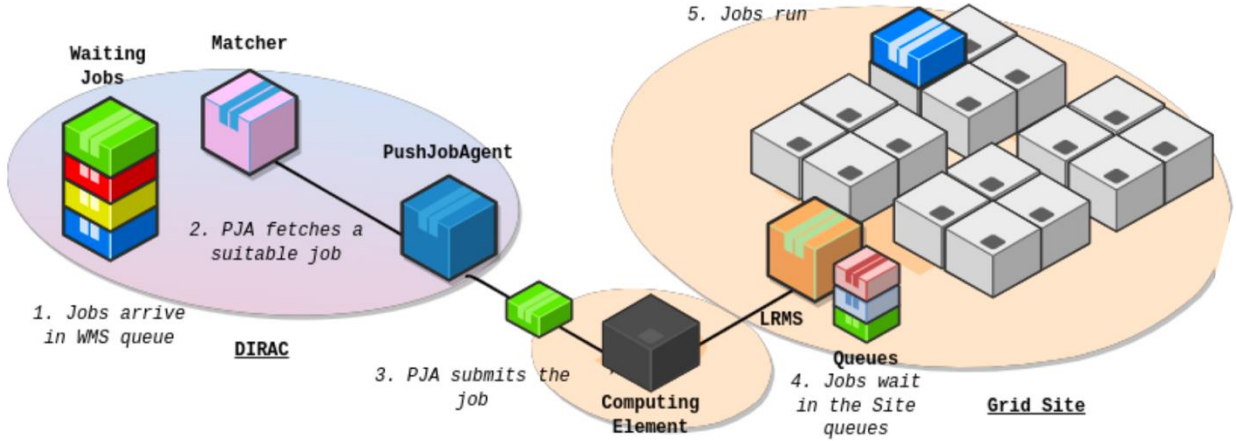


Solution for centers w/o external connectivity…

Use push model:
DIRAC's PushJobAgent

Submit jobs directly…

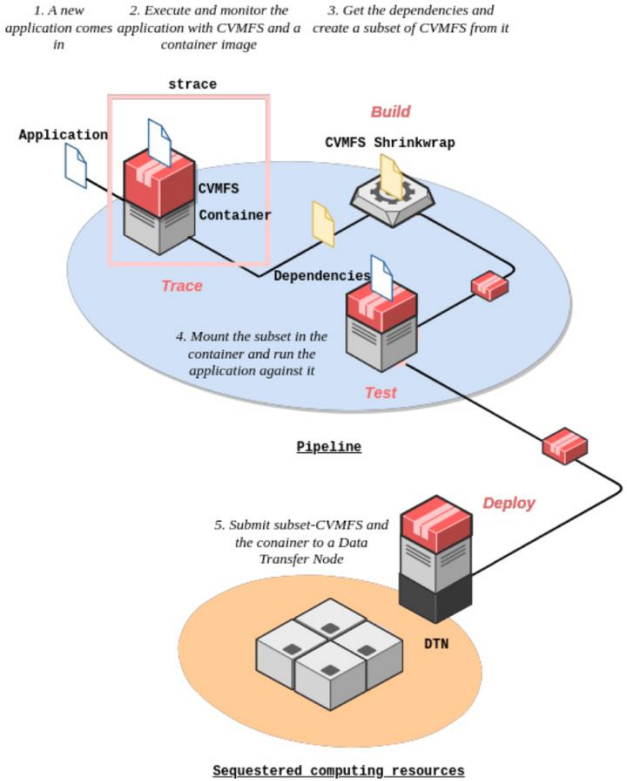
…handling interactions with external services before/after job execution
→ Jobs need to be adapted



But: scalability issues (RAM…)

…and CVMFS unavailability

- the *subset-cvmfs-builder* pipeline regularly extracts the needed dependencies and copies them on remote resources.



Conclusion

- HPCs are heterogeneous: no generic and unique solution
- Significant development needed with respect to the standard, grid-like solution

- Sites currently used: CSCS (CH), MareNostrum (BSC, ES), Kabre (CR)
- Allocations at SantosDumont (Petropolis, BR), Marconi100 (CINECA, IT), NERSC & OSC (US) in the past