ALICE, ATLAS, CMS AND LHCb REQUIREMENTS FOR EXPERIMENTAL PARTICLE PHYSICS COMPUTING RESOURCES

Concezio Bozzi, Erik Edelmann, Tomas Lindén, Gianfranco Sciacca, Mattias Wadenstein

Abstract

The members of the Lumi consortium are the Nordic countries and Belgium, the Czech Republic, Estonia, Poland and Switzerland. Experimental particle physics research in the LHC-experiments ALICE, ATLAS, CMS and LHCb is done in the Lumi member countries. This document collects the computing requirements and job profiles for these four LHC-experiments, so that those can be taken into account when configuring the Lumi super computer to enable it for this kind of research.

The LHC will be upgraded to the High-Luminosity LHC (HL-LHC) for Run 4. The increased luminosity of HL-LHC poses a significant computing challenge to the experiments. To meet this challenge the LHC-experiments are revising their computing models, optimizing their software and changing their analysis procedures in addition to looking for new sources of computing resources. Super computers like Lumi are very significant resources that could be enabled for experimental particle physics research [1].

Advances are being made in experimental particle physics to make use of GPU-resources in terms of machine learning, triggering and simulation, so the GPU-usage is expected to increase but currently the majority of computing is CPU-based.

As can be seen in the following the needs of the four different experiments are similar with CVMFS being used for software distribution and Singularity as the container technology. Jobs are submitted with ARC for ATLAS, CMS and LHCb and with ALiEn for ALICE.

1 Computing requirements

The ALICE, ATLAS, CMS and LHCb computing requirements and job profiles are summarized in the following two tables and the following text refers to the the tables. These numbers are also available on the VO card of each experiment [2, 3, 4, 5], but the ATLAS VO card is outdated.

A bandwidth of 80 Gbit/s from/to academic networks like NORDUnet/LHCOPN/LHCONE/GEANT is recommended for up to 10 kcores (peak capacity) [6].

The data is read directly from the Storage Elements by ALICE, CMS and LHCb so good bandwidth is needed in and out of the workernodes.

ATLAS only connects to external databases from the worker nodes with no data traffic directly from the nodes, so for ATLAS only low bandwidth outbound network traffic is needed.

The ATLAS bandwidth to the ARC server should be 40 Gbit/s for up to 10k cores (peak) matching throughput to/from the scratch area.

Table 1: ALICE, ATLAS, CMS and LHCb common computing requirements.

Service	ALICE	ATLAS	CMS	LHCb
Worker node (WN) network	outbound	outbound	outbound	outbound
Scratch storage	10 GB / job max local	25 GB / core max	20 GB / job max (local or cluster)	20 GB / job max
Job submission	ALiEn VOBox	ARC	ARC	ARC
Job accounting		ARC	ARC	ARC
Data caching		O(10TB) / 1000 cores optionally		
Software	CVMFS	CVMFS	CVMFS	CVMFS
Container technology	Singularity migrating to	Singularity	Singularity	Singularity
Calibration data		Frontier	Frontier	

Table 2: ALICE, ATLAS, CMS and LHCb job profiles.

Service	ALICE	ATLAS	CMS	LHCb
CPU	1 / multi core / full node	1 / n core no multi node jobs	8 / 1 core	1 core multi core coming no multi node jobs
RAM	2 GB min, 4 GB RAM + swap 6 GB	2 GB / core (can be less or more)	2 GB / thread	2 GB / thread 4 GB max non-swap
Runtime	< 30 h / 48 h max	< 24 / 36 h	48 h / 72 h max	queue limit

The LHCb amount of scratch storage shall be interpreted as 50 % each for downloaded input files and produced output files.

CVMFS is the CERN Virtual Machine File System which uses Squid Proxy [7]. It has been used to distribute software in the Finnish grid for many years. The CVMFS service can be shared across the experiments provided that the hardware is performant enough. The CVMFS mount points used by the experiments are preconfigured in the CVMFS packages.

CVMFS needs some kind of local cache on the workernodes. This could be on local disk of workernodes (if available) on /mnt/cache/cvmfs2. A reasonable minimum size for 4 VOs is probably 80 GB, but more is better. Another possibilty could be with the lower tier in RAM O(10 GB) and the higher tier in the scratch disk. For HPCs lacking local disk on nodes for cache a special configuration is used: preload all CVMFS repos from the 4 VOs on a shared O(TBs) cache area mounted on the nodes.

ATLAS needs Singularity with setuid root or enabling user namespaces and underlay. LHCb needs Singularity user namespaces.

The ATLAS and CMS Frontier service uses also the Squid Proxy and it can be shared with the CVMFS server.

The ALICE VOBox is a batch submission node with 1-2 cores, 4 GB RAM and 100 GB disk.

The LHCb 4 GB maximum amount of non-swap memory memory size is understood to be the virtual memory required per single process of a LHCb payload. Usually LHCb payloads consist of one "worker process", consuming the majority of memory, and several wrapper processes. The total amount of virtual memory for all wrapper processes accounts for 1 GB which needs to be added as a requirement to the memory size in case the virtual memory of the whole process tree is monitored.

References

[1] CMS Offline Software ; Computing, *HPC resources integration at CMS*, CMS-NOTE-2020-002 ; CERN-CMS-NOTE-2020-002, https://cds.cern.ch/record/2707936.

[2] ALICE VO card https://operations-portal.egi.eu/vo/view/voname/alice.

[3] ATLAS VO card https://operations-portal.egi.eu/vo/view/voname/atlas.

[4] CMS VO card https://operations-portal.egi.eu/vo/view/voname/cms.

[5] LHCb VO card https://operations-portal.egi.eu/vo/view/voname/lhcb.

[6] https://indico.cern.ch/event/770307/contributions/3312194/attachments/1808500/2952974/Networking_-_Best_Practice_for_Sites_and_Evolution.pdf

[7] http://cernvm.cern.ch/.