

The contribution of the WLCG Tier-2 site in Prague to the global WLCG operations

Dagmar Adamová,^{a,*} Alexandr Mikula,^b Martin Adam,^c Jiří Chudoba,^b Petr Vokáč^d
and Jana Uhlířová^b

^aNuclear Physics Institute of the Czech Academy of Sciences,
Rez 130, CZ 25068, Czech Republic

^bInstitute of Physics of the Czech Academy of Sciences,
Na Slovance 1999/2, Praha 8, CZ 18200, Czech Republic

^cCERN, European Organization for Nuclear Research
1211 Geneva 23, Switzerland

^dCzech Technical University,
Trojanova 13, 120 00 Praha, Czech Republic

E-mail: adamova@ujf.cas.cz

Ever since the start of the LHC operations in 2009, the mission of processing, management and storage of data collected by the LHC experiments has been executed by the Worldwide LHC Computing Grid (WLCG).

Building of the WLCG started at about 2001 and resulted in an extensive infrastructure of computing sites spread over 5 continents and interconnected with high capacity internet links. The current ecosystem comprises of ~1 million constantly occupied computer cores, ~1 ExaByte of storage and massive global networking with 10–100 Gb/s throughput. The computing centers involved in the WLCG are categorized as Tier 0, 1 and 2. Tier-0 is CERN, 14 Tier-1s are large computing centers providing computing power and disk and tape storage. Finally, there are 146 Tier-2 sites mostly universities and scientific institutes, which can store sufficient data and provide adequate computing power for simulation and analysis tasks. Tier-2s represent a very crucial part of the WLCG ecosystem delivering about half of the global WLCG resources.

In this contribution we will present an overview of operations of the Tier-2 site in Prague, Czech Republic, which provides computing and storage services for experiments ALICE and ATLAS and also some non-LHC experiments. This overview will show the crucial role of Tier-2 sites in the WLCG considering how many are involved. Our computing center is of a distributed character: the main part of resources is installed at the Institute of Physics (IoP) of the Czech Academy of Sciences (CAS); a part of the XRootD storage cluster for ALICE is operating at the Nuclear Physics Institute (NPI) of the CAS. Smaller clusters of the computing servers are located at the Faculty of Mathematics and Physics of the Charles University, and in CESNET, an association of the Czech Research and Educational institutions. We will provide a detailed overview of the Prague Tier-2 site infrastructure and operations with a special focus on the services provided for the ALICE experiment.

The Ninth Annual Conference on Large Hadron Collider Physics - LHCP2021

7-12 June 2021

Online

*Speaker

1. Introduction

In our age of extreme-scale computing the High Energy Physics (HEP), together with other data-intensive scientific communities, is crucially dependent on highly performant computing infrastructures. They are necessary for delivery of scientific results. One of the data driven research community involves the experiments at the Large Hadron Collider (LHC) [1] at CERN [2].

Ever since the start of the LHC operations in 2009 the mission of processing, management and storage of data collected by the LHC experiments has been performed by the Worldwide LHC Computing Grid (WLCG) [3]. The WLCG is the world's largest computing grid [4], see Figure 1. The computing centers involved in WLCG are ranked as Tier 0,1 and 2. Tier-0 is CERN, 14 Tier-1s are large computing centers providing computing power and disk and tape storage. Finally, there are 146 Tier-2 sites mostly at universities and scientific institutes, which can store sufficient data and provide adequate computing power. Tier-2s represent a very crucial part of the WLCG ecosystem delivering about half of the global distributed WLCG resources.

In this article we describe the global setup of WLCG and the important role of the Tier-2 sites in this ecosystem. As an example we present an overview of the Tier-2 site in Prague, Czech Republic, which provides computing and storage services for experiments ALICE [5] and ATLAS [6] @LHC. Special focus will be dedicated to the ALICE computing operations at this Tier-2 site.



Figure 1: Map of WLCG: 1 Tier-0, 14 Tier-1s and 146 Tier-2s.

2. The Worldwide LHC Computing Grid

The Worldwide LHC Computing Grid (WLCG) is a global computing infrastructure whose mission is to provide computing resources to store, distribute and analyze the data generated by the LHC, making the data equally available to all partners, regardless of their physical location [4]. When the building of the infrastructure started (~2001) the distributed or cloud computing was non-existent and the WLCG team had to invent all of the tools from scratch [7]. An extensive infrastructure of computing sites spread over 5 continents and interconnected with high capacity internet was gradually built and continuously upgraded until the current state-of-the-art ecosystem. It involves over 160 sites, running over 2 million tasks daily, with 1 million constantly occupied computer cores, 1 ExaByte of storage and network links with 10-100 Gb/s throughput.

3. praguecg2 : a distributed WLCG Tier-2 site in Czech Republic

praguecg2 [8] is one of the first computing centers included in WLCG. Its operations started already in 2004 and have been running steadily ever since. Currently praguecg2 is a modest size Tier-2 site with distributed resources. Its main part is installed at the Institute of Physics of the CAS (FZU) [9] in Prague and a substantial part of the storage cluster for ALICE is installed at the Nuclear Physics Institute of the CAS (NPI) [10]. There are also some resources provided by other institutions in Prague and at the city of Ostrava.

The complete stack of resources involves ~10,000 CPUs, 6.65 PetaBytes of disk space and network connectivity of the range between 10 Gb/s and 100 Gb/s. Very important is the 100 Gb/s connection to the WLCG global network infrastructure LHC Open Network Environment (LHCONE) [11]. The compute, storage and data management services provided by praguecg2 are used predominantly by ATLAS and ALICE@LHC, see Figure 2.

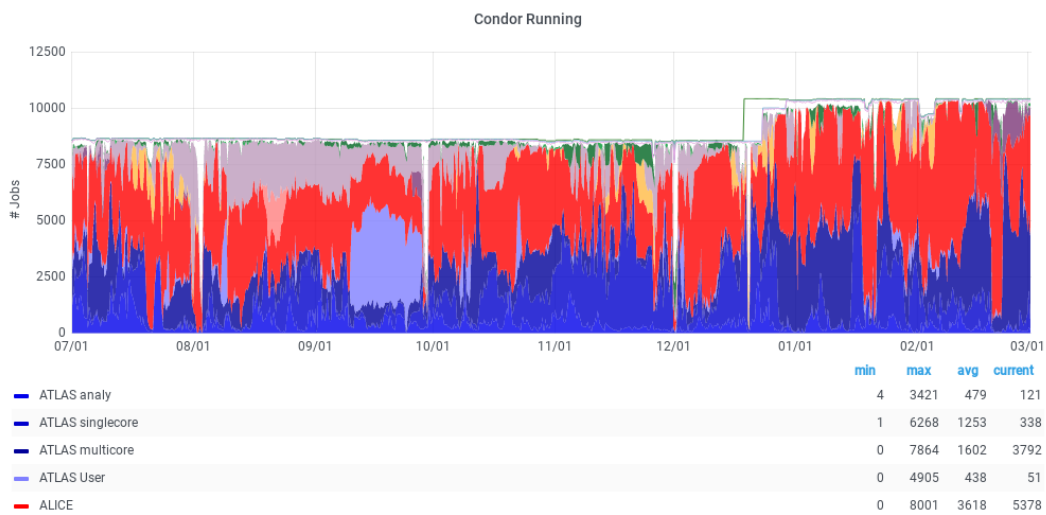


Figure 2: Running jobs profile for recent 9 months provided by the on-site monitoring . Main CPU consumers are ALICE and ATLAS.

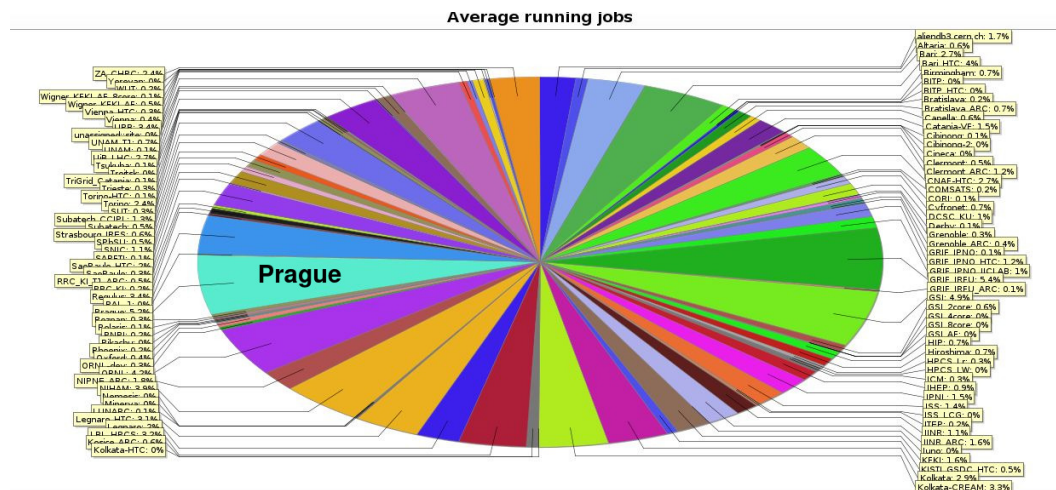


Figure 3: Delivery of ALICE compute resources: Tier-2s share during 3/2020–4/2021. praguecg2: 5.2%.

4. ALICE operations at praguecg2

ALICE computing is special due to the obligation of each participating institution to deliver compute and storage capacities based on ALICE M&O-A. Czech ALICE group is quite small in comparison with the local ATLAS group and this translates also into a smaller share in available stack of resources. Despite this disadvantage we manage to deliver required capacities, see Figure 3. This is achieved thanks to praguecg2 site ability to operate almost without interruptions and longer downtimes and a suitable fairshare set at the main site at FZU which allows the ALICE jobs to use almost all CPUs that are not taken by other experiments jobs. The ALICE jobs are very efficient in using any opportunistic resources. As a result, millions of ALICE jobs are processed at praguecg2 each year (e.g. 16,5M in 2015–2019). Our XRootD ALICE storage holds $\sim 47,799,700$ data files and ~ 67 PB of data was downloaded from our storage during 3/2020–4/2021.

5. Summary

The WLCG is the most sophisticated data-taking and analysis system ever built for science, providing near real-time access to LHC data [4]. In its hierarchical structure the 146 Tier-2 centers represent irreplaceable part of the WLCG resources, as explained earlier. As an example of the WLCG Tier-2 site we presented the praguecg2 center in the Czech Republic. Its contribution to the WLCG operations is clearly recognized and will be growing in the coming years. We also presented how the Prague site contributes to the computing operations of the ALICE experiment, in particular showing that the site complies with the ALICE obligatory resource delivery requirements.

Acknowledgments

This work was supported by the Ministry of Education, Youth and Sports of the Czech Republic project LTT17018.

References

- [1] The Large Hadron Collider; <https://home.cern/science/accelerators/large-hadron-collider>
- [2] CERN portal; <https://home.cern/>
- [3] Ian Bird et al: Update of the Computing Models of the WLCG and the LHC Experiments; <https://cds.cern.ch/record/1695401/files/LCG-TDR-002.pdf>
- [4] Worldwide LHC Computing Grid; <https://wlcg-public.web.cern.ch/>
- [5] The ALICE Collaboration: The ALICE experiment at the CERN LHC; 2008 JINST 3 S08002
- [6] The ATLAS Collaboration: The ATLAS Experiment at the CERN Large Hadron Collider; 2008 JINST 3 S08003
- [7] Ian Bird: WLCG Challenges and Collaboration; <https://indico.cern.ch/event/828520/contributions/3534877/attachments/1968040/3272985/WLCG-Challenges-LHCOPN-130120.pdf>
- [8] Martin Adam et al: Distributed resources of Czech WLCG Tier-2 center; EPJ Web of Conferences 245, 03034 (2020)
- [9] Institute of Physics CAS; <http://www.fzu.cz/en/>
- [10] Nuclear Physics Institute CAS; <http://www.ujf.cas.cz/en/>
- [11] LHCONE; <https://lhcone.web.cern.ch/>