A Cloud Demonstrator for Belle II, T2K and Hyper-K

Silvio Pardi[1*], Sophie King[2], Mathieu Guigue[3], Aurélien Bailly-Reyre[3], Marko Bracko[4]

[1] Istituto Nazionale di Fisica Nucleare (INFN) Sezione di Napoli, Napoli, Italy.
[2] King's College London (KCL), London, UK.
[3] Laboratoire de Physique NucléAire et des Hautes éNergies (LPNHE), Sorbonne Université, Université Paris Cité, CNRS/IN2P3, Paris, France.
[4] Institut Jožef Stefan, Ljubljana, Slovenia.

* Corresponding author. Email: spardi@na.infn.it, sophie.king@kcl.ac.uk, mguigue@lpnhe.in2p3.fr, abaillyr@lpnhe.in2p3.fr, marko.bracko@ijs.si
Manuscript submitted June 17, 2022; revised July 15, 2022; accepted September 27, 2022.
doi: 10.17706/ijapm.2023.13.1.1-9

Abstract: The international physics community has spent large efforts in the last decade to utilize different types of resources in their computing model, in order to increase the available computing power and involve more people and institutions in the activities. In this paper we present the work done to define and implement a set of common tools for integrating cloud computing resources into the computing framework of the Belle II, T2K and Hyper-Kamiokande (Hyper-K) experiments to produce Monte Carlo simulations and run user analyses. Starting with the virtual machine lifecycle tool VCYCLE, we create a custom setup able to support and manage the environment of the three experiments, then integrate the machinery into the DIRAC infrastructure of Belle II and in the GridPP DIRAC service used by T2K and Hyper-K. In order to expand the number of usable cloud infrastructures, we extended the software capability by integrating token based authentication in the Openstack module of VCYCLE. With this modification we are able to run Monte Carlo production jobs on the European Grid Infrastructure (EGI) Federated Infrastructure demonstrating the flexibility of implemented the solution.

Key words: Cloud, high energy physics, Monte Carlo simulations.

1. Introduction

Belle II [1], T2K [2] and Hyper-Kamiokande (Hyper-K) [3] are three large experiments that have experimental facilities located in Japan, working on the exploration of particle physics, searching for signals of new physics within flavor physics and neutrino physics respectively.

The three collaborations are working together in the context of project JENNIFER2 [4] funded under the Horizon2020 program of the European Union as a Marie Sklodowska Curie Action of the RISE program, under grant n.822070. Within the JENNIFER 2 project, Work Package 5 (WP5) is dedicated to computing and common techniques, in particular, Task 5.1 is devoted to exploiting computing and data handling technologies for the three experiments. The main goal is to study a set of common tools for the computing model of the Belle II, T2K and Hyper-K experiments, which users from Japanese and European Research centers can use in a transparent way. The main outcome of the task is the creation of a demonstrator.

In this work we present the technologies and the specific setup used by the working group to create a common platform, integrating several cloud infrastructures provided by the involved institutes, within the computing framework of the three experiments. This is done in the context of DIRAC [5]; a software framework for distributed computing. We describe our customized setup of the VCYCLE [6] software, designed to be used in a multi-experiment environment. In addition, we show how we adapted the main tool in order to use token based authentication and expand the demonstrator over the European Grid Infrastructure (EGI) [7] Federated Cloud infrastructure.

The paper is organized as follows. In Section 2, we present the computing model of each of the three experiments, describing the frameworks and adopted tools. In Section 3, we describe the technological choices for the demonstrator and the specific setup implemented. In Section 4, we show the early results obtained by using the cloud resources through the platform. Finally, in Section 5, we summarize our work and discuss the conclusions.

2. Computing Model Analysis

In order to define a set of common tools, the computing models of the three experiments were first thoroughly examined as follow:

2.1. Towards a Distributed Computing Model for Hyper-K

The Tokai-to-Kamioka (T2K) experiment is a long-baseline neutrino oscillation experiment located in Japan. It has been taking data since 2010 and provides world-leading measurements of the neutrino oscillation parameters, as well as neutrino interaction measurements and new physics searches. The T2K collaboration is formed of around 500 members from 70 institutes in 12 countries. Hyper-Kamiokande (Hyper-K) is the successor to T2K, with an improved neutrino beam, a new intermediate detector and a far detector with a fiducial volume that is eight times larger. Hyper-K aims at pinning down CP-violation phenomena in the lepton sector, and has a wide programme of physics including nucleon decay searches, astrophysics and beyond the standard model physics. There are currently around 450 members within the Hyper-K collaboration, from 93 institutes across 19 countries. Construction began in 2021 and Hyper-K will begin taking data in 2027. The T2K computing model centers around the near detector, ND280, and is largely based on the Large Hadron Collider (LHC) Grid infrastructure. The T2K virtual organization (VO) uses the multi-VO GridPP [8], [9] instance of DIRAC, hosted at Imperial College London. The GridPP DIRAC instance is provided as a service to experiments, with the development being done by the GridPP team, largely driven by the needs of LHCb and without contribution from the experiments using the framework. T2K has computing resources linked to DIRAC at multiple UK Grid sites, L'Institut national de physique nucléaire et de physique des particules (IN2P3) (France), Simon Fraser University (SFU) (Canada) and High Energy Accelerator Research Organization (KEK) (Japan). The Hyper-K computing model is currently very similar to that of T2K, but with plans to add more sites for both job submission and data storage across the globe, including Italy, Poland, Mexico and Switzerland. When T2K ends, the resources currently allocated to ND280 through the T2K VO, and the data that is stored, will be transferred to the Hyper-K VO. A summary of the three core components, job management, data storage, and software distribution are presented below:

2.1.1. Job management

T2K and Hyper-K jobs are run by using singularity sandboxes distributed on the EGI Cern Virtual Machine File System (CVMFS) [10], with RAL hosting the Stratum-0 server. Jobs submitted to LHC Grid resources are managed through the GridPP instance of DIRAC, and mainly concern Monte Carlo production and calibration tasks. T2K and Hper-K also have resources (CPU and GPU) allocated in Kamioka (Japan), on the Cedar cluster of Compute Canada and at smaller sites around the world. At the moment these are handled by local batch systems and do not connect to DIRAC. Going forward, Hyper-K is looking for ways to connect more of these resources to the centrally managed DIRAC system.

2.1.2. Data storage

T2K and Hyper-K use DIRAC for basic data management through the Dirac File Catalogue (DFC) for data stored on LHC Grid resources. For Hyper-K, raw and processed data, MC production and calibration files from the near, intermediate, and far detectors will be stored on the LHC Grid resources. KEK CC acts as the Tier-0 (T0) for ND280, i.e the primary storage of raw data, where data is stored on a disk and tape combination High Performance Storage System (HPSS). Once Hyper-K begins taking data, KEK will also provide the T0 storage for the intermediate detector, and the lab in Kamioka will become a T0 for the far detector. The T0 data is independent of the LHC Grid; however, there is a Grid storage element (SE) at KEK CC where data is transferred and then synced to the T1 sites SEs. All raw data is stored at the T0 site and on two T1 SEs, and MC production is stored at one T1 site and one T2 site. The definition of T1 and T2 is borrowed from the LHC definition, which has limits on the expected down-time of storage and the availability of technical on-site staff. In some cases, we may slightly relax the rules; so, a site that is a T2 by the LHC definition, could in practice be a T1 for T2K/Hyper-K if deemed suitable. Hyper-K storage of raw and processed data and MC production after 10 years of data taking is expected to be of the order of tens of PBs. Both T2K and Hyper-K also have additional storage managed by Integrated Rule-Oriented Data System (iRODS) (hosted in Japan) and Nextcloud (hosted in the UK).

2.1.3. Software versioning and distribution

All ND280 software for T2K is version controlled using the T2K GitLab hosted in Poland. The GitLab continuous integration features are utilized, with 'Runners' being executed with a Docker-in-Docker setup. Built images for each software tag are uploaded to a T2K repository. The software versioning and distribution system for Hyper-K is still under development, but it will continue to be centered around Git and the use of containers.

2.2. The Belle II Computing Model

Belle II is an international collaboration aimed at studying Flavour Physics using the data collected at the Super KEKB e+e- collider at the KEK laboratory located in Tsukuba (Japan). Belle II involves 126 Institutes spread over 26 countries/regions for a total of more than 1000 scientists. Belle II will search for new physics beyond the Standard Model by accumulating 50 times more data (~50ab-1) than its predecessor Belle which operated at the KEKB accelerator up to 2010. The experiment is in the data taking stage, and from 2019 new raw data files are produced and collected in some of the main sites of the collaboration. The computing model [11] has been designed to accomplish all the tasks needed to process and reprocess RAW data, producing Monte Carlo simulation, and smoothly run Physics analysis over a distributed infrastructure.

The data center participating in the activities are classified in three main classes:

- RAW Data Centers: responsible for the processing and reprocessing of raw data and archiving of data on tape for long term data custodially. They participate in MC production. A full copy of raw data is stored at KEK while the second copy is distributed in different countries with the following share: USA (30%), Italy (20%), Germany (20%), Canada (15%) and France (15%).
- Regional Data Centers: responsible for maintaining the processed data in the so-called mini Data-Summary Table (mDST) format, obtained from raw data reconstructions, and from MC in which skimming is done. Regional Data Centers participate in MC production as well.
- MC production Center: this class of sites, includes all the resources that provide computing power for Monte Carlo Production, including opportunistic resources.

The collaboration takes advantage of more than 55 sites distributed all over the world, offering pledged and opportunistic resources with different technologies, including cloud resources.

The Belle II workload management system used by Belle II is based on the DIRAC framework empowered with BelleDIRAC extensions, created to support the specific workflows. DIRAC is configured with a distributed setup, with the main servers hosted at KEK and a set of slave machines in different countries. The system integrates heterogeneous resources with different interfaces.

The applicative software is distributed over the computing nodes via The CernVM File System (CVMFS), also recently used for the distribution of the pilot code. As regarding the data handling, recently a RUCIO [12] instance has been adapted and integrated into the DIRAC framework and is now used for all data management activities, while AMGA [13] and FTS3 [14] are used respectively as the metadata catalogue and data movement tool.

2.3. Common Tools

From the analysis of the computing models and tools used by the participating experiments, we individuated the following components as pillars of the JENNIFER2 Cloud demonstrator:

- CentOS: A common operating system to be deployed on the different clouds.
- CVMFS: A mountable filesystem used to distribute applicative software for the VMs.
- DIRAC: A common framework of technologies that provide a workload management system.

For a Cloud management system, we identify VCYCLE as the first candidate to form the basis of a common infrastructure. This solution, developed as part of the GridPP Cloud/VM efforts, offers a full set of tools for VM lifecycle management, and is already designed to be integrated in the DIRAC framework. VCYCLE implements the so-called vacuum model on IaaS Cloud services such as OpenStack which is one of the most popular cloud technologies used by the scientific community. In addition, a previous positive evaluation was done within the Belle II experiment, guaranteeing the basic know-how to simplify the full adoption.

3. The Cloud Infrastructure

3.1. VCYCLE Deployment

VCYCLE has been implemented in a server running in the Belle II site of Napoli, in Italy. The basic setup is composed of two web-endpoints, the first used to distribute VM contextualization scripts and the second to receive log files coming from the VMs instantiated over the different cloud infrastructures. This is depicted in Fig. 1. The third component is the VCYCLE engine which is responsible for the instantiation of VMs over the available Cloud infrastructures, and monitoring and managing the full life cycle of machines, up to the shutdown signal to be sent when the VM completes the assigned task or when something goes wrong during contextualization.

In order to use VCYCLE for the three experiments, we had to develop a multi-vo configuration able to instantiate VMs suitable for the specific exigence of T2K, Hyper-K and Belle II. More specifically, taking advantage of the flexibility of VCYCLE, for each Cloud infrastructure available we configure a separate configuration file in /etc/vcycle.d/. Each file contains the credential to access at the relative cloud and three different Virtual Machine profiles, in order to instantiate VMs with different contextualization instructions.

The golden image is based on CERNVM latest available version, while the contextualization is done via cloud-init and the user-data script downloaded by the VM at the first boot. The user-data script is responsible for performing the basic configuration steps, configuring CVMFS and starting the pilot job, configured thanks to a json file, called pilot.json, which contains the specific configuration of the DIRAC server which should be contacted, and the information on the sites, and VO. In each cloud infrastructure used for the demonstrator, we implemented a cache service based on squid, over a separate VM, that is also referenced in the VCYCLE configuration file of the specific endpoint.

The testbed was set up to use the Cloud infrastructure of INFN-Napoli, then a demonstrator has been extended to the LAL and LPNHE infrastructure, all based on Openstack offering a standard authentication system based on the Openstack keystone services. From the DIRAC point of view, each Cloud is configured as a separate site with prefix VCYCLE in the Belle II DIRAC system, and with prefix CLOUD in the GridPP DIRAC system that serves T2K and Hyper-K.

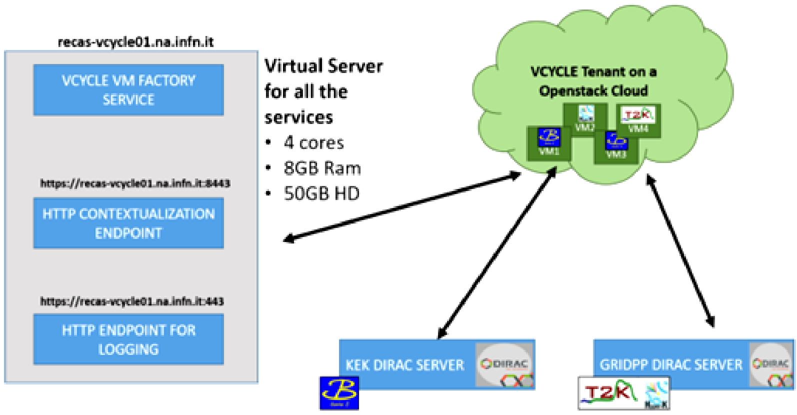

Fig. 1. The basic setup of VCYCLE. The two webservers and the VCYCLE engine are all running on a single VM.

3.2. Integration of EGI Federated Cloud

The European Grid Infrastructure (EGI) coordinates a distributed cloud system called EGI Federated Cloud [15]. Such international infrastructure integrates multiple endpoints, running over different institutes in several European countries. The core part of the EGI Federated Cloud is represented by an authentication and authorization infrastructure (AAI), called EGI-Checkin [16], fully based on token technology. In order to expand our demonstrator over the EGI Federated Cloud, and help the three collaborations to take advantage of additional resources, we first had to empower the VCYCLE code by developing a new module for the token based authentication. In particular, on the openstack_api.py module we defined the following new variables:

egi_checkin_url: URL of EGI Check-in authentication endpoint
egi_checkin_client_id: the single ID on the federated cloud
egi_checkin_client_secret: the client secret
egi_checkin_refresh_token: A long term token to generate an access token on the fly
egi_project_id: the Project ID assigned for a specific Cloud Endpoint

We then added the code displayed in Fig. 2 for the authentication. In the case that egi_project is defined, we first refresh the access token from the EGI Check-in service, then ask for a scoped token to the specific endpoint that should trust the specific VO.

```
if self.egi_project_id:
      vcycle.vacutils.logLine('Auth with EGI Token')
      access_token = refresh_access_token(self.egi_checkin_client_id,self.egi_checkin_client_secret,self.egi_checkin_refresh_token,self.egi_checkin_url)
      vcycle.vacutils.logLine('EGI Access token created')
      ep = find_endpoint("org.openstack.nova", site=self.egi_site).pop()
      os_auth_url = ep[2]
      self.egi_token, _ = get_scoped_token(os_auth_url, access_token, self.egi_project_id)
      vcycle.vacutils.logLine('EGI scoped token created' + ' project  ' + self.egi_project_id )
      jsonRequest = { "auth": { "identity": { "methods" : ["token"],
                                              "token": { "id": self.egi_token }},
                                "scope": {"project": {"id": self.egi_project_id }}
                               }
                    }
```

Fig. 2. Implementation of token based authentication in VCYCLE.

To use this authentication method, it is necessary to create a user community within EGI Federated Cloud, and then propagate the basic VM image on the infrastructure supporting the specific community. In the context of our demonstrator, we previously tested this new approach for Belle II VO using CESGA and INFN-Catania resources for the basic setup. Finally, a full collaboration has been established with the IN2P3-IRES institute by signing a specific SLA that allowed them to stably integrate their resources via the EGI Federated cloud.

4. Demonstrator in Action

After the setup of the cloud endpoints in both DIRAC infrastructures, we were immediately able to instantiate VMs and run production jobs for the three experiments. Regarding Belle II, we have 4 sites currently offering resources via JENNIFER2 tools: VCYCLE.Napoli.it, VCYCLE.LAL.it, VCYCLE.LPNHE.fr and VCYCLE.EGI.eu. Since the start of 2020 JFY in April an integrated CPU power of 39.3kHS06 has been provided for the collaboration. Cloud resources have been mainly used for MC production, however around 12% of the CPU time has been for user activities. In Fig. 3 we show the details through the graph produced by the accounting system of Belle II production DIRAC.

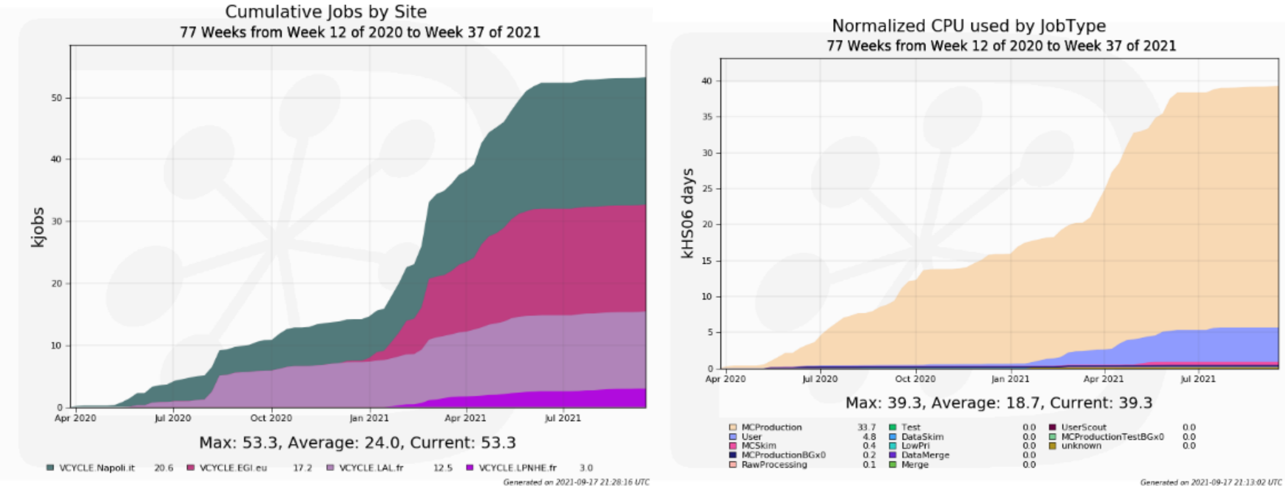

Fig. 3. Belle II activities from the accounting system of Belle II production DIRAC.

For T2K and Hyper-K, we have 2 sites currently configured on the GridPP DIRAC system, for both the VOs, which are: CLOUD.RECAS-NAPOLI.it and CLOUD.GRIF.fr. Since the start of 2020 JFY the opportunistic resources have processed a total of 14kJobs. As the graph in Fig. 4 shows, the largest part of the activities has been run by Hyper-K, participating in Monte Carlo campaigns.

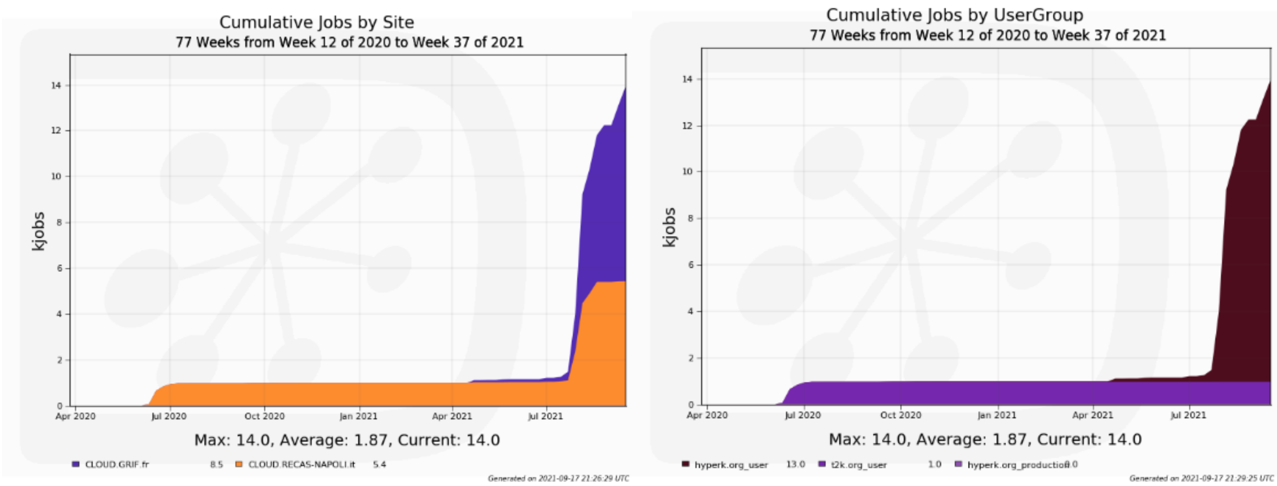

Fig. 4. T2K and Hyper-K activities from the accounting system of GridPP DIRAC.

5. Conclusions

A first version of a common cloud infrastructure for the three experiments, T2K, Hyper-K and Belle II has been released. This enables them to integrate into their respective computing infrastructures, new cloud resources using a set of common tools and a shared Virtual Machine Manager system hosted in Napoli. In addition, the development of the new authentication interface for VCYCLE enables it to expand the demonstrator over the EGI Federated Cloud, increasing the amount of opportunistic resources available for the collaborations. This is the result of synergies created thanks to the JENNIFER2 initiative. The activities will continue in the coming years, working to expand the demonstrator over additional cloud endpoints and studying how to exploit new technologies and how to integrate cloud storage space in a multi-VO environment.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

All authors have participated in designing the model and setup different aspects of the testbed. SP supervised the research. SP and SK wrote the paper; All authors agree with the final version of the manuscript.

Funding

This work has been supported by funds from the European Union's Horizon 2020 Research and Innovation Programme, MSCA-RISE-2018 call, project JENNIFER2, GA 822070.

Acknowledgment

The authors would like to acknowledge the support provided by the GridPP Collaboration, in particular the GridPP DIRAC developers at Imperial College London and the KEK members who support the integration with Belle II DIRAC infrastructure. Thank you to EGI computing infrastructure for the support in integrating the Federated Cloud and to all sites that contributed stable resources, INFN-Napoli, IN2P3-IRES, IN2P3-LAL and IN2P3-LPHNE and to sites that contributed to testing the EGI infrastructure, INFN-Catania and CESGA.

References

[1] Abe, T., Adachi, I., Adamczyk, K., *et al*. (2010). Belle II Technical Design Report. Retrieved from https://arxiv.org/abs/1011.0352

[2] Abe, K., Abgrall, N., Aihara, H., *et al*. (2011). The T2K experiment. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 659(1),* 106–135.

[3] Abe, K., Aihara, H., Aimi, A., *et al*. (2018). Hyper-Kamiokande Design Report. Retrieved from https://arxiv.org/abs/1805.04163

[4] JENNIFE2 Project. An MSCA-RISE project funded by EU under grant n.822070. Retrieved from http://www.jennifer2-project.eu

[5] Tsaregorodtsev, A., Bargiotti, M., Brook, N., *et al*. (2008). DIRAC: A community grid solution. *J. Phys.: Conf. Ser.*, *119*, 062048.

[6] McNab, A., Love, P., & MacMahon, E. (2015). Managing virtual machines with Vac and Vcycle. *J. Phys.: Conf. Ser., 664*, 022031.

[7] The European Grid Infrastructure. Retrieved from http://www.egi.eu
[8] Bauer, D., Colling, D., Currie, R., *et al*. (2015). The GridPP DIRAC project - DIRAC for non-LHC communities. *J. Phys.: Conf. Ser., 664*, 062036.
[9] Britton, D., Cass, A. J., Clarke, P. E. L., *et al.* (2009). GridPP: The UK grid for particle physics. *Phil. Trans. R. Soc. A, 367*, 2447-2457.
[10] CVMFS. Retrieved from https://cernvm.cern.ch/fs
[11] Pardi, S., Nardoab, G., Russo, G., *et al.* (2016). Computing at Belle II. *Nucl. Part. Phys. Proc., 273-275*, 950-956.
[12] Serfon, C., Mashinistov, R., Stefano, J. S., Villanueva, M. H., Ito, H., Kato, Y., Laycock, P., Miyake, H., & Ueda, I. (2021). Integration of RUCIO in Belle II. *Proceedings of the 25th International Conference on Computing in High Energy and Nuclear Physics (CHEP 2021): Vol. 251*, 02057.
[13] Kwak, J. H., Park, G., Huh, T., Hwang, S., *et al*. (2015). Improvement of AMGA python client library for Belle II experiment. *J. Phys. Conf. Ser., 664*, 042041.
[14] Ayllon, A. A., *et al*. (2014). FTS3: New data movement service for WLCG. *J. Phys.: Conf. Ser., 513*, 032081.
[15] Castillo, E. F., Scardaci, D., & Garcia, A. L. (2015). *Procedia Computer Science, 68*, 196-205.
[16] EGI-CheckIn project. Retrieved from https://wiki.egi.eu/wiki/AAI

Silvio Pardi is a senior computer science technologist at INFN. Since 2002, he has been contributing significantly to the development of the Italian Grid Infrastructure. He his scientific responsible of networking for the IBISCO Infrastructure in Napoli for HPC and HTC applications. He is coauthor of more than 200 articles in international journal and conferences. Currently is Infrastructure coordinator for Belle II international Experiment.

Sophie King is a research associate working on the T2K and Hyper-K experiments; currently at King's College London and previously at Queen Mary University of London. Her research focuses on neutrino interactions with matter, how these interactions are measured with detectors, and the inherent systematic uncertainty associated with modelling this. She also works with grid and cloud computing, primarily for the purpose of Monte Carlo simulation of neutrino interactions within our detectors.

Mathieu Guigue is researcher in experimental neutrino physics Laboratoire de physique nucléaire et des hautes énergies (LPNHE), Sorbonne Université, Université Paris Cité, CNRS/IN2P3. His research interests are neutrino physics whitin the T2K and Hyper-K experiments.

Aurélien Bailly-Reyre is a computer science enginieer at Laboratoire de Physique nucléAire et des Hautes éNergies (LPNHE), Sorbonne Université, Université Paris Cité, CNRS/IN2P3. He has busy as cloud and grid administrator. He also has experience in high performance computing technologies.

Marco Bracko is an assistant professor at University of Maribor, Faculty of Chemistry and Chemical Engineering; Jožef Stefan Institute (JSI). He works on experimental high energy physics in this context he is member of Belle and Belle II collaborations.