Computing

4th JENNIFER2 General Meeting
3 April 2025
Dr. Silvio Pardi on the behalf of Computing Group

EU grant n.822070

JENNIFER2 WP5 Computing and common techniques

- **Task 5.1: Common Computing and data handling**
- Task 5.2: Common DAQ and remote controls issues
- Task 5.3: Statistical methods for analysis combination
- Task 5.4: Generators and phenomenology

5.1 Deliverable: “Common Cloud Computing Demonstrator”

Key people involved: S.King, S.Pardi, M.Bracko, T.Kuhr, M. Guigue with the contribution of many other people

1° Computing WorkShop - Task 5.1

CERN 12 December 2019
10 participants in person and 3 participants via remote connection

- 6 presentations from Jennifer2 members
- 2 invited speakers
- Final Working Session

Very nice atmosphere

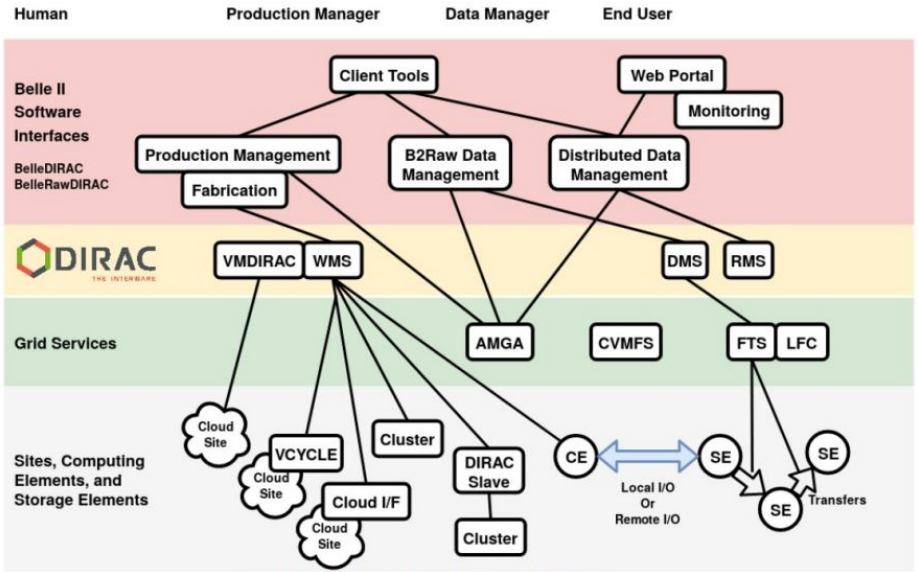
https://agenda.infn.it/event/20616/

Belle II Computing Model

Framework:
DIRAC + BELLE2DIRAC Extension
GRID, Cloud, SSH Cluster
Transformation system

Tools:
CVMFS for software distribution
GIT+bitbucket
AMGA
RUCIO (ongoing)

Experiment status:
Data Taking started
MC Production
Skimming
User Analysis
RAW Data: 1 copy in KEK + 1 copy distributed
Estimated RAW Data size: 12 PB/year at max. luminosity



M. Bračko, Jennifer2 CompWS, CERN, 2019/12/12

T2K - HK Computing Model

Currently T2K and HK are separate experiments.

Framework:

DIRAC provided by GRIDPP with no specific customization.

DFC for Data Management

Tools:

CVMFS for software distribution

GITLAB

CMAKE

Containers

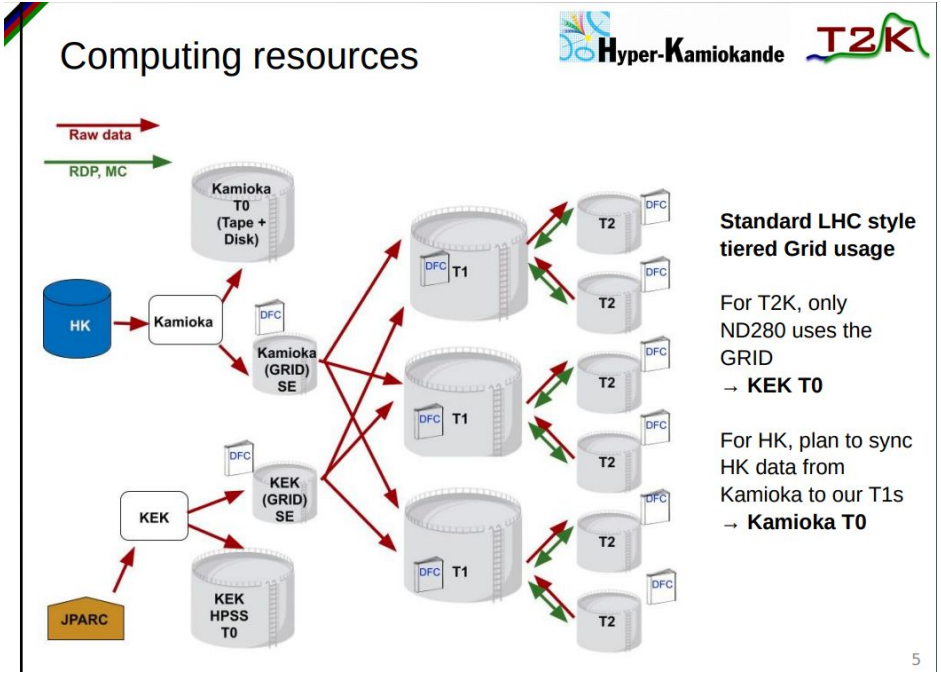
Experiment status:

T2K

- In data taking
- MC Campaign

HK

- Estimated data rate: 5 TB of RAW Data per day



Demonstrator Startup

As outcome of the event we decided to setup a common Cloud Computing infrastructure for the three experiments based on a set of technologies of common interest that included:

- DIRAC as workload management system
- VCYCLE as Virtual Machine Manager
- Openstack as Cloud IaaS
- CVMFS for software distribution
- CentOS as Operative system

VCYCLE for JENNIFER2 demonstrator

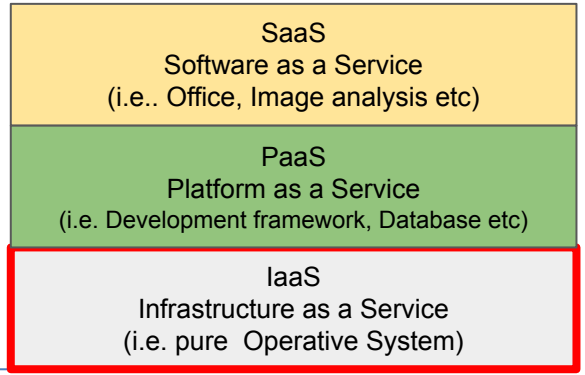
VCYLCE is VM lifecycle manager developed by GRIDPP, it is designed to create VMs on Cloud endpoints offering EC2, Openstack or Azure interface.

VCYCLE can be easily integrated in DIRAC and the accounting system is compliant with APEL.

VCYCLE has been selected to be used as interface

for the JENNIFER2 demonstrator

https://www.gridpp.ac.uk/vcycle/

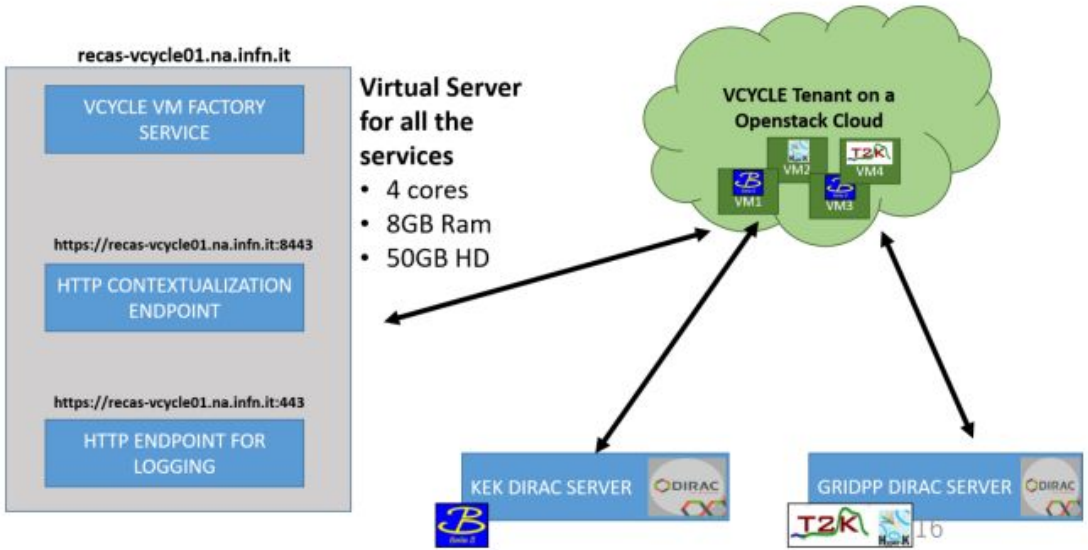


JENNIFER 2 Cloud Demonstrator

For the Jennifer2 demonstrator we created a single VCYCLE service infrastructure and we attached it to Different Openstack Clouds using a standard local account:

- LAL
- LPNHE-GRIF
- Napoli

We setup two profiles one for Belle II DIRAC, and one for T2K and HyperK DIRAC



EGI Federated Cloud

In order to expand the number of resources that the two community can use, we exploit the possibility to use the Federated Cloud of EGI (The European Grid Infrastructure)

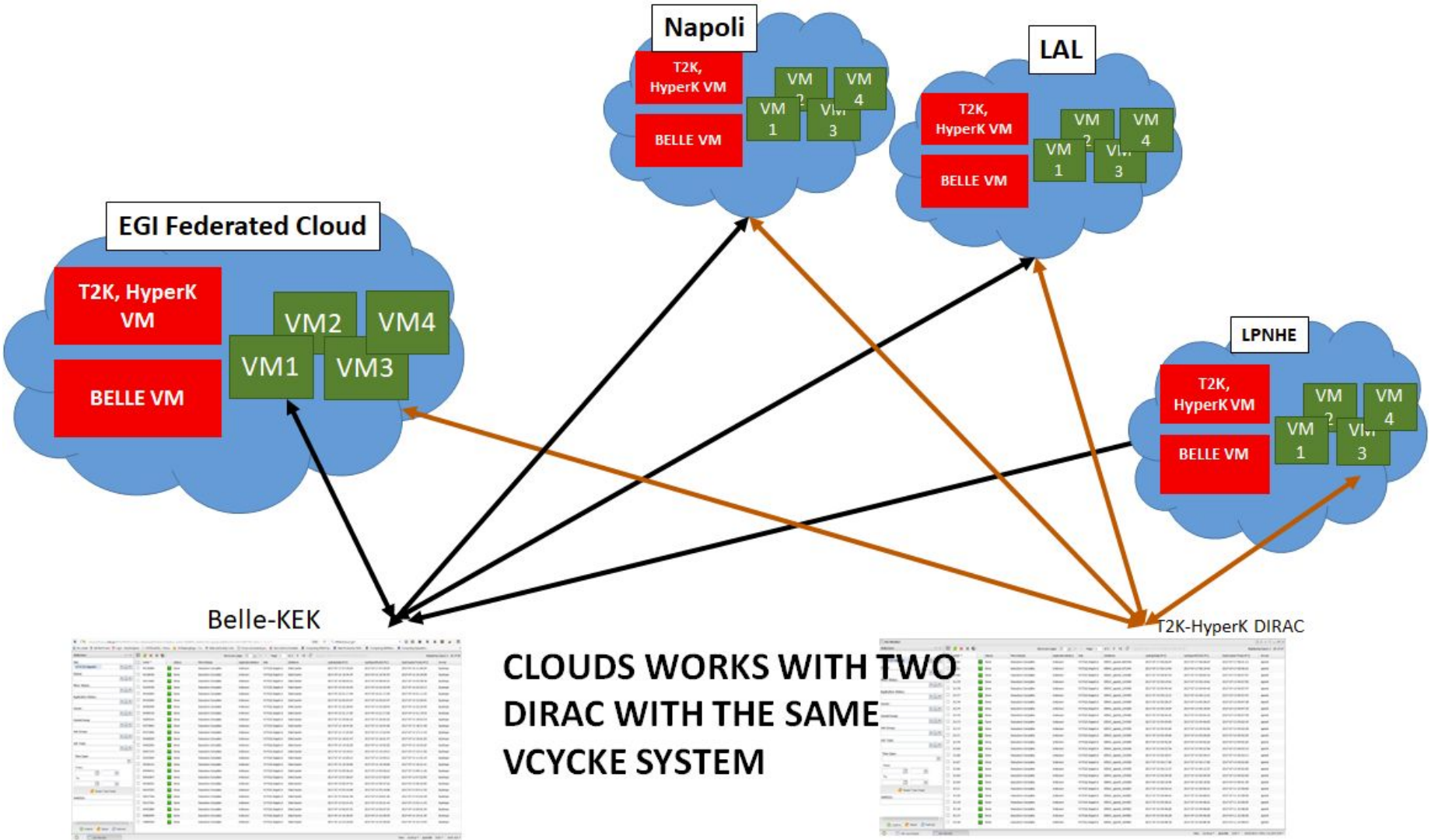
It consists of a set of Cloud Endpoints distributed in several European Countries, glued together with the EGI Federation Tools.

Add EGI Federated Cloud Resources

VCYCLE has been expanded with a new authentication method. In the EGI testing environment we have access to three Openstack endpoints which are:

- CESGA
- IFCA LCG2
- INFN Catania

After a stress test phase, a stable cloud infrastructure has been integrated in the Jennifer2 demonstrator, provided by IN2P3-IRES institute which dedicated a set of resources for the project.



CLOUDS WORKS WITH TWO DIRAC WITH THE SAME VCYCKE SYSTEM

Cloud Demonstrator

EGI Conference:
https://indico.egi.eu/event/5000/contributions/14307

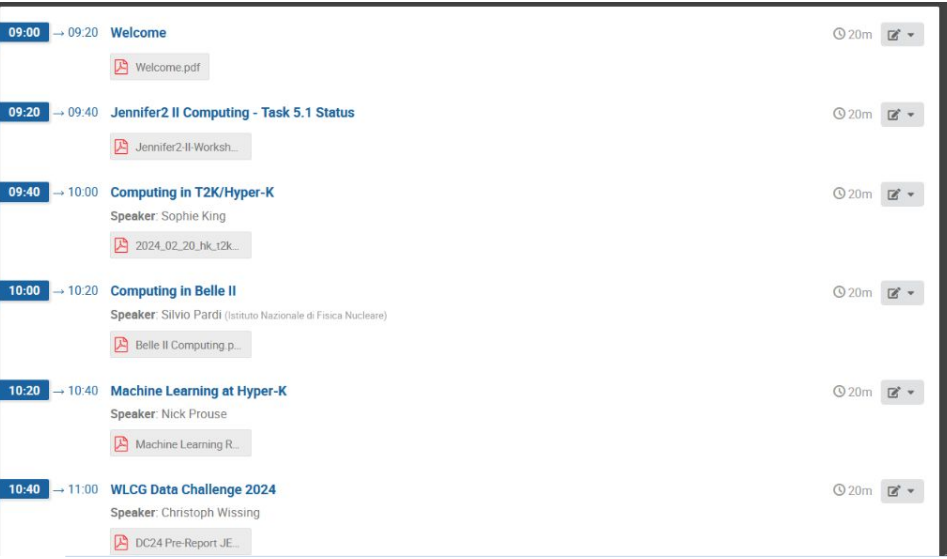
The 11th International Conference on Engineering Mathematics and Physics (ICEMP22): http://www.icemp.org/

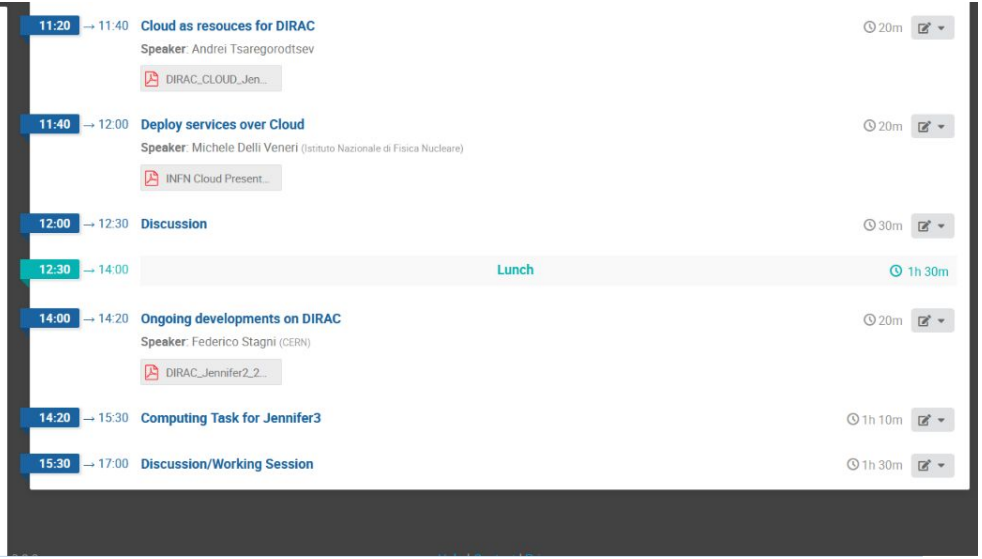
- Best presentation awarded
- Proceeding on *International Journal of Applied Physics and Mathematics*

2° Computing WorkShop - Task 5.1

Organized at CERN the 20 February 2024 https://agenda.infn.it/event/39895/

4 Invited speakers, discussion on the achieved results and we ideas for JENNIFER3





Milestones for JENNIFER3

Two main topics for Task 5.1 and Task 5.2

- Cloud Technologies
- Network Data Challenge

Milestones

D.5.1 Organization of a workshop on the usage of cloud computing technologies in the context of Belle II, T2K and Hyper-K (month 24)

D.5.2 Report on the joint Data Challenge Session testing the network links from Japan to the main EU Data Centres involved in Belle II and Hyper-K (month 36)

Key People: Sophie King, Silvio Pardi, Nick Prouse, Mathieu Guigue, Marcko Bracko, Cristiano Bozza, Thomas Kuhr

Task 5.1 Cloud Technologies

Investigate Cloud Computing for Analysis and MC

Study how to exploit the Cloud technologies by integrating resources, in a stable way, into the computing model of three experiments, Belle II, T2K and Hyper-K.

This can be done through the implementation of a common interface such as the DIRAC Cloud Computing Element, for integration of virtual machines into the MC factory and for testing the PaaS/SaaS services offered by INFN and CNRS Clouds.

Cloud Technologies

The usage of Cloud for Code Testing/Development or Analysis is another hot topic in high energy physics.

A discussion could be open on the possible usage of PaaS/SaaS service.

Some interesting technologies: JupyterHub, Dask, Kubernetes, Cluster service on demand, distributed database application.

Possibility to make synergy with other tasks of WP5

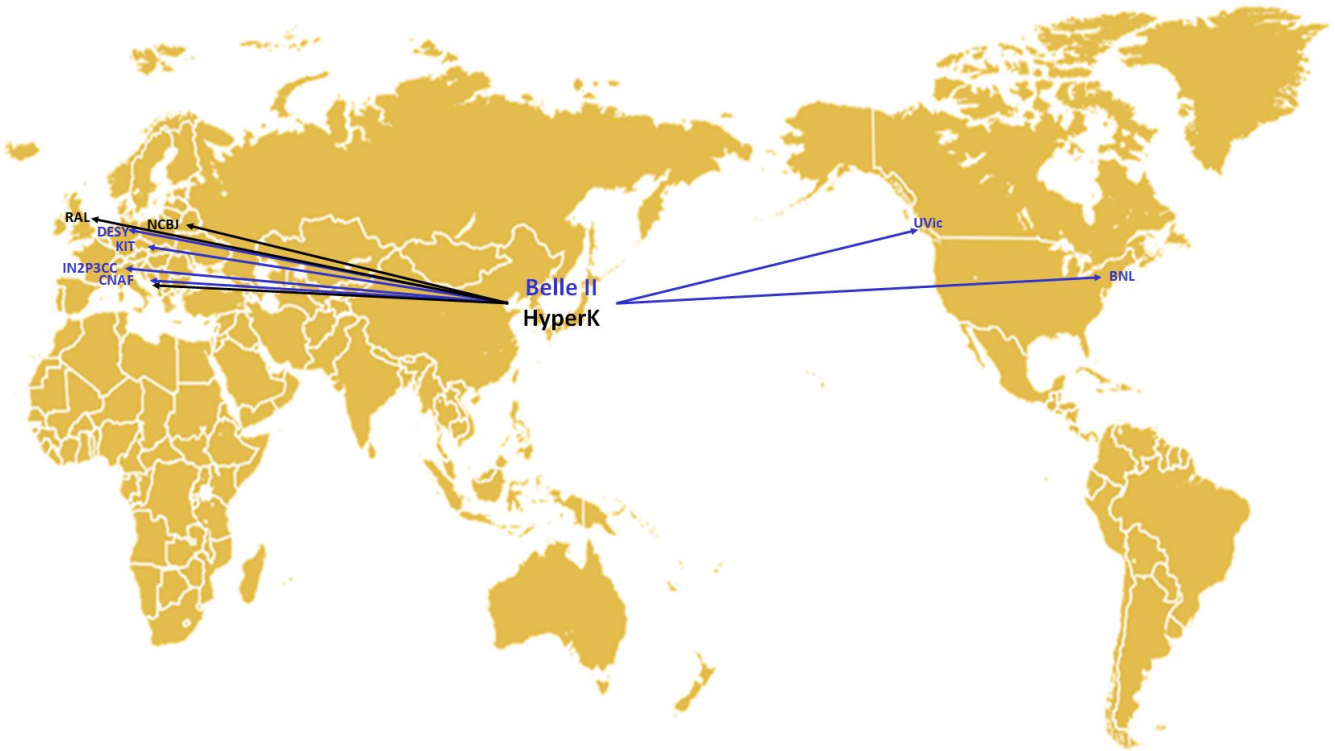
Task 5.2 Networking

A Joint Network Data Challenge:

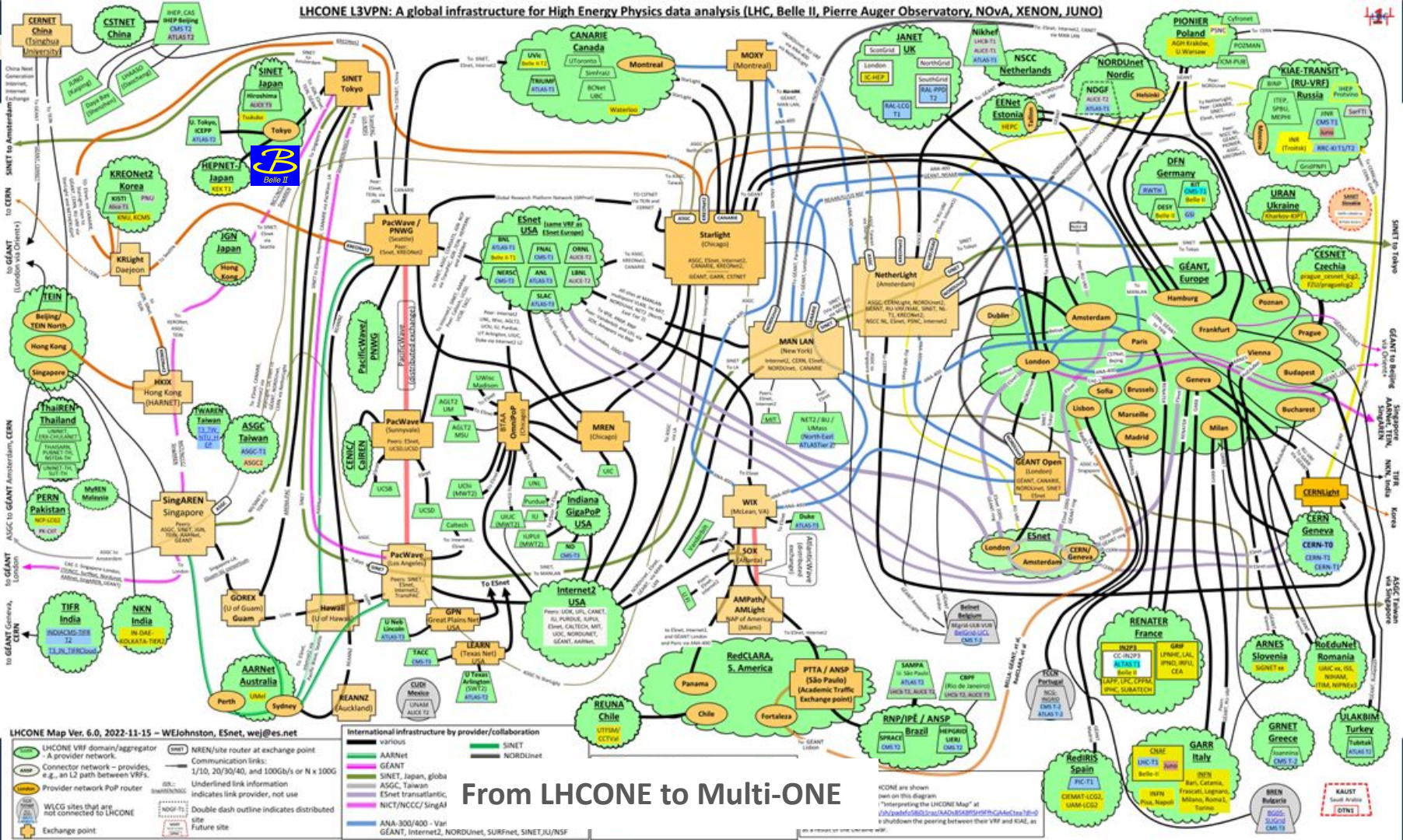
The goal of this task is to test the speed of the international networks links connecting Japan vs EU institute, assess the efficiency and stability of the data management tools. It will be done by performing a set of joint stress tests sending data from storage to storage from the sites in Japan to the European-based Belle II and Hyper-K data centres such as CNAF, CC-IN2P3, DESY, KIT, SIGNET, NCBJ, Napoli and RAL.

A set of common tools will be used to monitor the network links and the performance will be compared with the estimated requirements of the two experiments.

Belle II and Hyper-K in DC27



The idea is to organize a joint Belle II/Hyper-K data challenge session and integrate these tests within the context of WLCG DC27. This will involve stressing the specific network links connecting Belle II and Hyper-K sites to Japan, concurrently with the traffic generated by LHC experiments.


LHCONE L3VPN: A global infrastructure for High Energy Physics data analysis (LHC, Belle II, Pierre Auger Observatory, NOvA, XENON, JUNO)
From LHCONE to Multi-ONE
LHCONE Map Ver. 6.0, 2022-11-15 – WEJohnston, ESnet, wej@es.net
International infrastructure by provider/collaboration

WLCG Next Data Challenge 2027

The New WLCG Data challenge Campaign is under organization by WLCG

As Hyper-K is working to define which storages can be involved in the test

As Belle II already joined the DC24, for the DC27 will define new additional paths to test and will take the opportunity to test token based authentication.

Plans for Task 5.1 and Task 5.2

Kick-off meeting for the computing activities, probably in remote.

For the first milestone we will start the activities on cloud, also we will select a set of topics to bring in the Workshop and start to think in which place organize the event.

For the Data Challenge we will start to involve people, define the sites and the resources to be used and start to figure out how to organize a set of pre-test of incremental difficulty.

THANK YOU FOR YOUR ATTENTION

Backup

WLCG Data Challenge

- WLCG has mandated to execute data challenges (DC) for HL-LHC
  - Demonstrate readiness for expected HL-LHC data rates by a series of challenges
  - Increasing volume/rates
  - Increasing complexity (e.g. additional technology)
  - A data challenge roughly every two years
- DOMA is the coordination and execution platform
  - Data Organization Management & Access
    - Forum across all LHC experiments to address **technical** challenges
  - DC coordination across the LHC experiments and beyond
    - Suited dates
    - Reasonable targets
    - Functionalities
  - Help in orchestration
- No pressure on sites to increase their capacity
  - But can we improve the existing infrastructure?

Year	% of HL-LHC
2021	10%
2024	25%
2027	50%
2028?	100%

Belle II Data Challenge 2024

What should be exercised during DC24:

Technology that can be stressed: Network, DDM, FTS, Storages, Monitoring System, Protocols, IAM

Main goal: Emulate data transfer conditions in a Belle II high-lumi scenario

Our current estimation for such scenario is 40 TB per day.

Transfers from KEK to RAW Data Centers according to our distribution schema (30%BNL, 20%CNAF, 15% IN2P3CC, 15%UVic, 10%DESY, 10%KIT)

Considering that the average speed needed to transfer 40TB/day is 3.7Gbit/s in outbound at KEK vs all the Raw Data Centers.

- Min - The target speed to achieve is 3x3.7Gbit/s = **11.1 Gbit/s**
- Max - The target speed to achieve is 5x3.7Gbit/s = **18.5 Gbit/s**

Workshop Outcome

New approach to integrate Cloud resources into the experimental framework. CloudCE

New ideas to exploit Cloud Infrastructure for user analysis or prototyping

Discussion on Machine Learning concepts for Hyper-K

Strengthening the cooperation between the two computing groups

Increasing synergy with the Data Organisation, Management, and Access (DOMA) working group of WLCG in preparation for the next joint Network Data Challenge in 2026

Conclusions

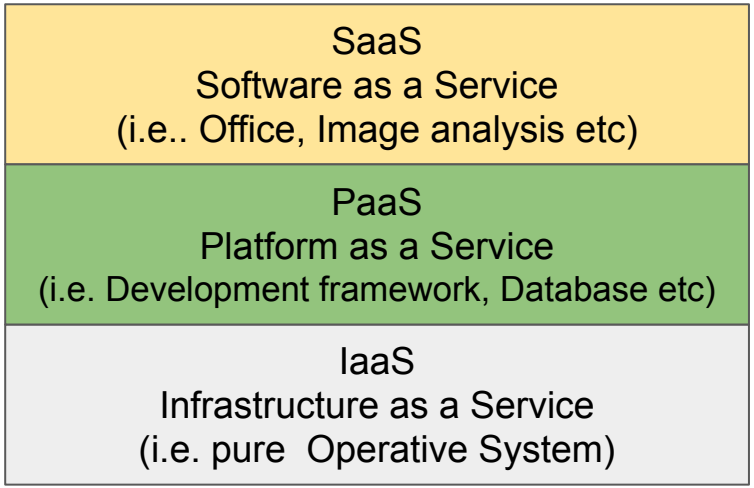
The collaboration between the three experiments—T2K, Hyper-K, and Belle II—has been very successful, and the milestone of task 5.1 has been achieved.

This achievement enables the integration of new cloud resources into their respective computing infrastructures using a set of common tools and a shared Virtual Machine Manager system hosted in Naples.

These results are the product of synergies created through the JENNIFER2 initiative and will serve as the foundation for the upcoming JENNIFER3 initiative.

Cloud Computing

Cloud Computing is a technologies for resource provisioning under the paradigm of virtualization.



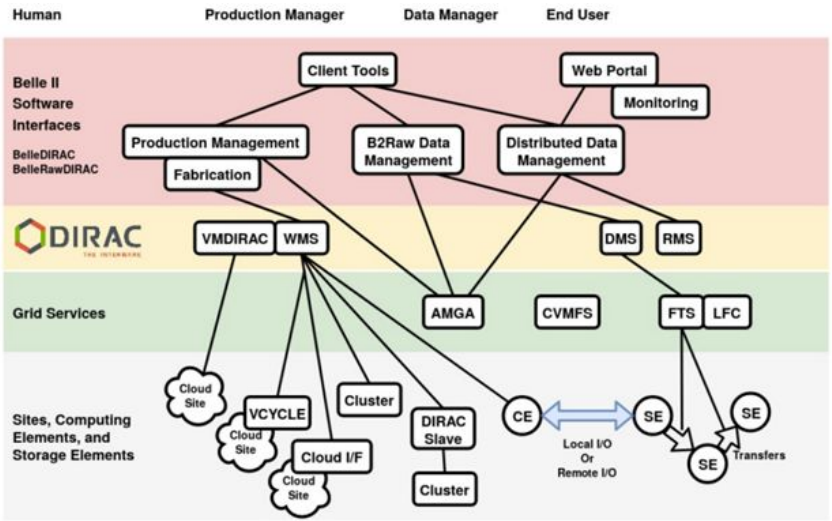
https://csrc.nist.gov/publications/detail/sp/800-145/final

DIRAC Framework for Belle II, T2K, HyperK

DIRAC is a framework for data and workload management. It enable users to submit jobs and retrive data over different computing resources distributed everywhere.
The three experiments Belle II, T2K, HyperK use DIRAC to perform MonteCarlo simulations, analysis, skimming over the GRID.

Through DIRAC is possible to send job vs several kind of resources via GRID interface, SSH, and cloud as well.

DIRAC
THE INTERWARE



M. Bračko, Jennifer2 CompWS, CERN, 2019/12/12