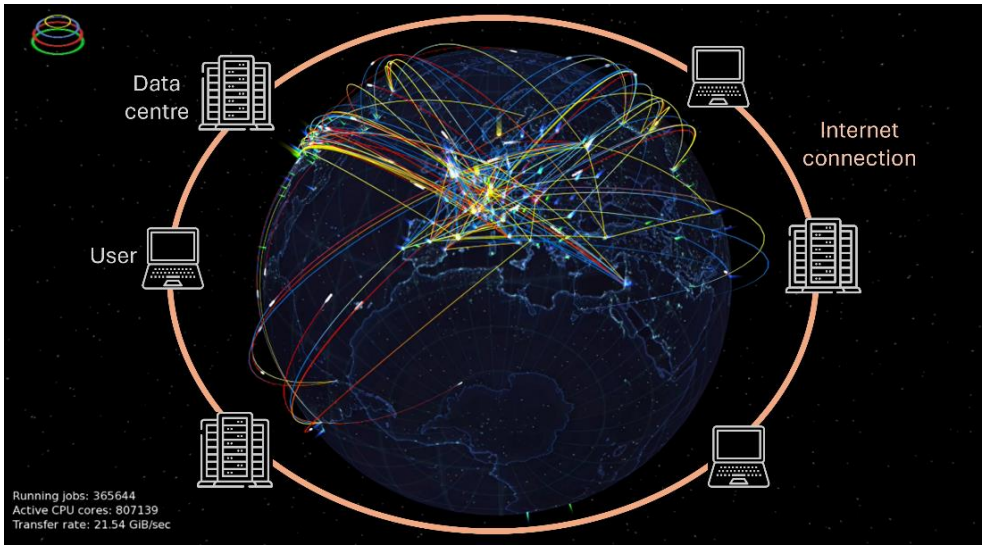
The Largest Computer in the World

Physicists are big IT consumers. They use it to analyse, simulate or store their data. Long-term data storing is essential to physics research, as is the ability to access it from institutes all over the world, since a same collision will be scrutinised and analysed many times by several teams around the world.



Credit: CERN, adapted

What is a computing grid?

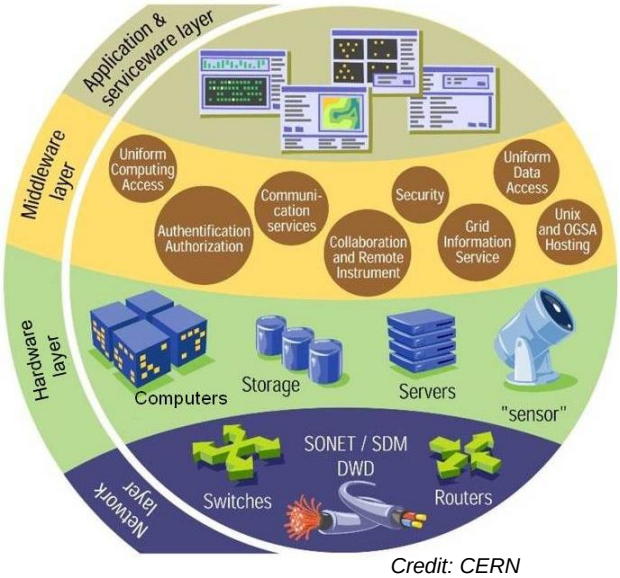
This image illustrates the Worldwide LHC Computing Grid activity captured live in August 2017. The computing grid connects computers located on every continent, regardless of their architecture or type. Thanks to the advances in data transfer, distances are abolished and all these resources can be used for a common purpose.

Since the 1960s, physicists have been using what are known as data centres. In most cases, the main centre is located close to the experiment generating the data. Its role is to store the data and make it available via catalogues. Researchers also run monitoring tasks and a first level of processing, called data preparation, which enables them to check the quality of the data, apply initial corrections and intervene rapidly if needed. This is followed by other tasks: simulation, reconstruction, reformatting and filtering, and training machine learning algorithms. Today's experiments require so much computing and storing power that a single centre is no longer sufficient. It has become necessary to network hundreds of geographically dispersed data centres and pool their resources.

This is known as the Worldwide LHC Computing Grid (WLCG), which enables researchers to connect to remote servers via their own computers using broadband networks. In practical terms, when a physicist submits an application to the WLCG, the system redirects it to a data centre that has the resources to process the request. That is why this technology is called the grid, an allusion to the electric power grid, which allows any user to obtain electrical power through a simple interface (the socket) without having to think about where the energy comes from or how it was produced.

The WLCG developed by particle physics computer engineers makes it possible to harness the computing power (processors, memory, etc.) of thousands of computers, while benefitting from their individual characteristics to meet the specific and varied needs of the field. The tasks demanding very long execution times – for instance, the detailed simulation of the products of LHC collisions and the detectors' response – are handled by the data centre network of the community, but also by High Performance Computers (HPCs) provided by other research areas interested in the technical aspects. On the other hand, data processing – or reprocessing, if improvement is needed – campaigns require extensive access to storage space; they are therefore reserved for centres that have a copy of the data and enough storage capacity to make the results available. Finally, each institute has a certain amount of autonomy in choosing the tasks that serve their teams' research themes. Monitoring this entire system increasingly involves artificial intelligence, particularly for error detection.

The WLCG's architecture can be represented in layers, each with a distinct function: firstly, the applications, i.e. the software used by users. Then, the middleware, a set of software installed on all the machines which enables the various elements to work together; it also decides which calculations to be performed and which data is to be stored where. Then, the resources – data centre computers, storage systems, data catalogues, etc. Lastly, the computer network (the Internet), which ensures connectivity between the resources on the grid, just as the electric power grid connects power plants and sockets.



Credit: CERN

Drawing on the experience and lessons learned from a decade of operating the LHC, the technical teams are working on designing the next-generation grid to meet expected internal needs as well as those of other fields, for example through the European Open Science Cloud (EOSC) project.

A layered architecture

The computing grid is organised in layers, each with a distinct function.

Translated from 'Le plus grand ordinateur du monde',* Passeport pour les deux infinis, *IN2P3/Dunod, https://pass2i.ijclab.in2p3.fr/ (French); updated with ICHEP2024 conference (https://ichep2024.org) for IPPOG