20 May 2022

Status of the Computing for Research in Africa

Ghita Rahal

CC-IN2P3, CNRS, France

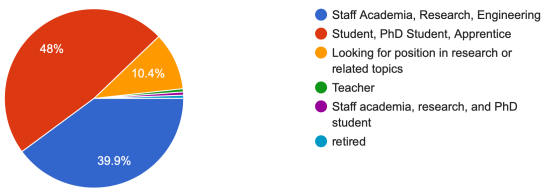
Introduction

Research in any Science needs nowadays strong computing services to extract results and make discoveries.
What we define as computing services might rank from the underlying structure, namely networks, computers, storage, to the applications and the software but as well new techniques such as Artificial Intelligence to extract the expected results.

In order to estimate the overall needs in this field, we have launched a survey including all the people that we could reach, participants in ASFAP as well as the attendants to the 2021 ACP conference that was held in mars 2022.
In this paper we summarize the information that was gathered in the different questions of the survey and extract some general observations. Possible guidelines to improve the situation are drawn in the conclusion.

1. Panel distribution

173 people filled in the survey out of which 165 were African citizens. 26 countries of the African continent were represented.
82.4% of the African citizens are based in Africa, the rest is what is defined as the diaspora, i.e., people that are based in other continents.
For the Non-African citizens, the motivation for participation was mainly an already established collaboration with African colleagues or students.



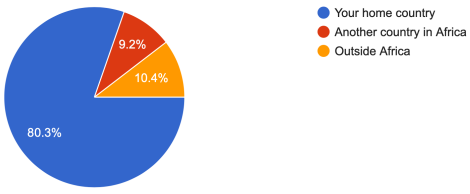


Figure 1: Left: Position of the participants. Right: Place where activity is performed.

Figure 1 highlights the positions of the participants: 48% are students and 39.9% hold a position in academia, research, engineering. More than 89% work in Africa, whether it is in their own country or in another. 10.4% are abroad.

2. Field of Research

The field of research in which the participants are working is spread among many disciplines as shown in the figure below where the contributions with 1 person where excluded for the lisibility of the picture.

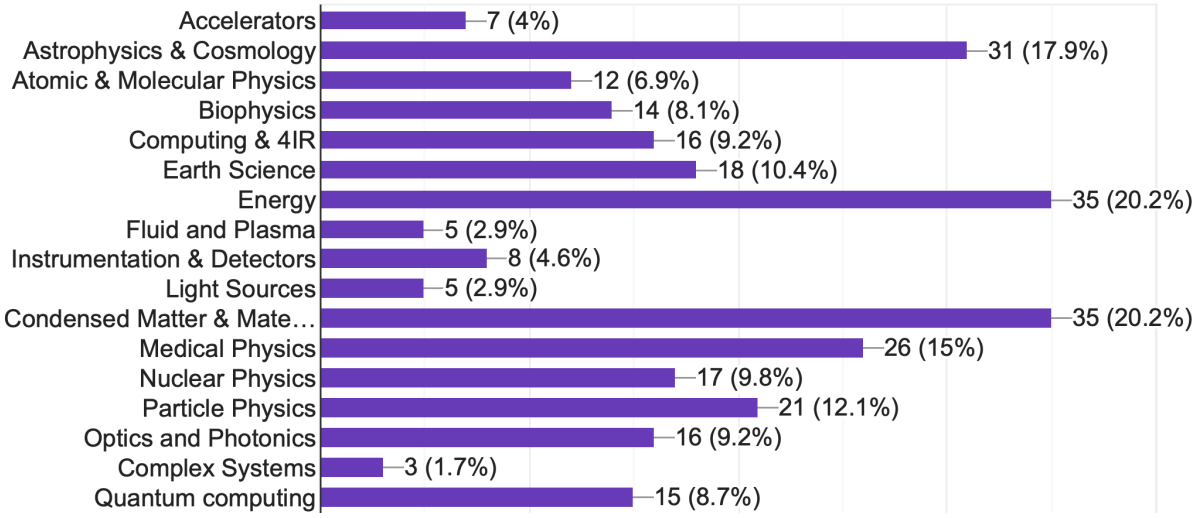


Figure 2: Field of Research covered by the participants

3. Properties of the exploited data

Depending on whether the scientist is working in collaboration or alone, the origin and volume of the data that is analysed in order to extract results varies a lot. Figure 3 below shows the distribution with also the size of the sample

- 36% of the participants use data from their experiment and 31% use their own data.
- The majority of the samples are at the Gigabyte and Megabyte level. It is to be noted that the samples of Terabytes- and Petabytes-level are stored internationally.
- Another graph not shown here shows that 76% of the scientists store and exploit their data in their own laboratory or in their own country, while 19.7% work at the level of international collaborations.

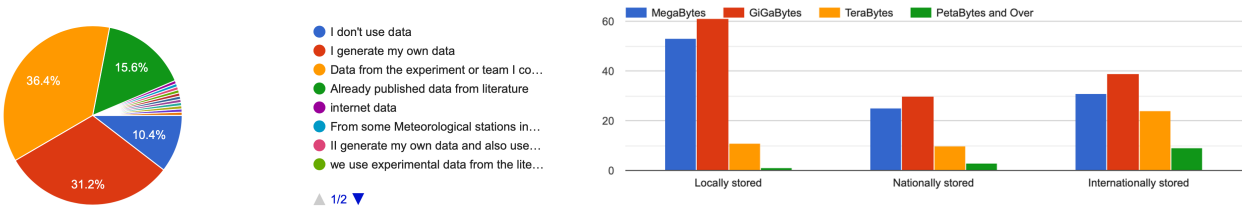


Figure 3: Left: shows where the exploited data comes from and Right: size of the sample for each storage location

4. Software and Tools to exploit data

Figure 4 illustrate the type of software that scientists are using:

- In the left graph of fig 4, we observe that 52.9% of the scientists use collaboration software and 48.4% use commercial software to exploit their data. Only 24.6% of them use only their own software. This might be due to various reasons that we have not studied in this survey.
- On the right pie, usage of Artificial Intelligence (AI) or Deep Learning (DL) is represented. Already 20.9% of the scientists have introduced it in their toolbox.
  What is striking is that more than 73% of the scientists would like to use it but are prevented to do so either because they cannot find the information and/or training they need, or because they lack computing resources to execute it. This is important to keep in mind because of the growing importance of AI in many fields of sciences.

Figure 4: Left distribution of the type of software that is used by the researchers. Right: Usage of Artificial Intelligence or Deep Learning in the research.

5. Status of the infrastructure and tools

We have questioned the participants about the resources or knowledge that appears to be necessary to be able to use computing to extract research results. The different points that have been evaluated are:

- Access to data and access to information: about half of the people experience blocking points leading to insufficient access to the data.

- Hardware resources, Computing power, Storage space and Networks: 55 to 60% of the participant find it insufficient. When it comes specifically to local infrastructure, 66% of the people find it insufficient. The reason of the difference between the 2 percentage above is certainly that some people use international infrastructures that are more efficient: 20% of the people are based outside their home country and about 40% claim that they use resources abroad (see fig.1).
- Mentoring: training, guidance, lectures, etc..: 54% of the participants don't find it sufficient.

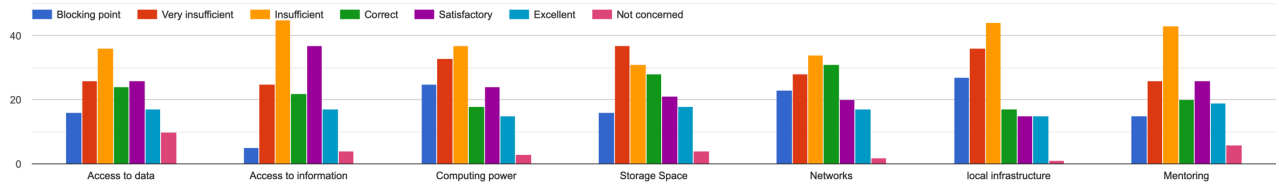


Figure 5: Main blocking or satisfactory points to be able to exploit data: the 5 first graphs show the quality of the different resources such as information or computing power. The 6th graph shows graphically the feeling of the user about their local infrastructure. The 7th and last graph is about the quality of the mmentoring.

6. Main identified bottlenecks

In Figure 6, we specifically question the users about the main bottlenecks there are facing. 3 main points appear:

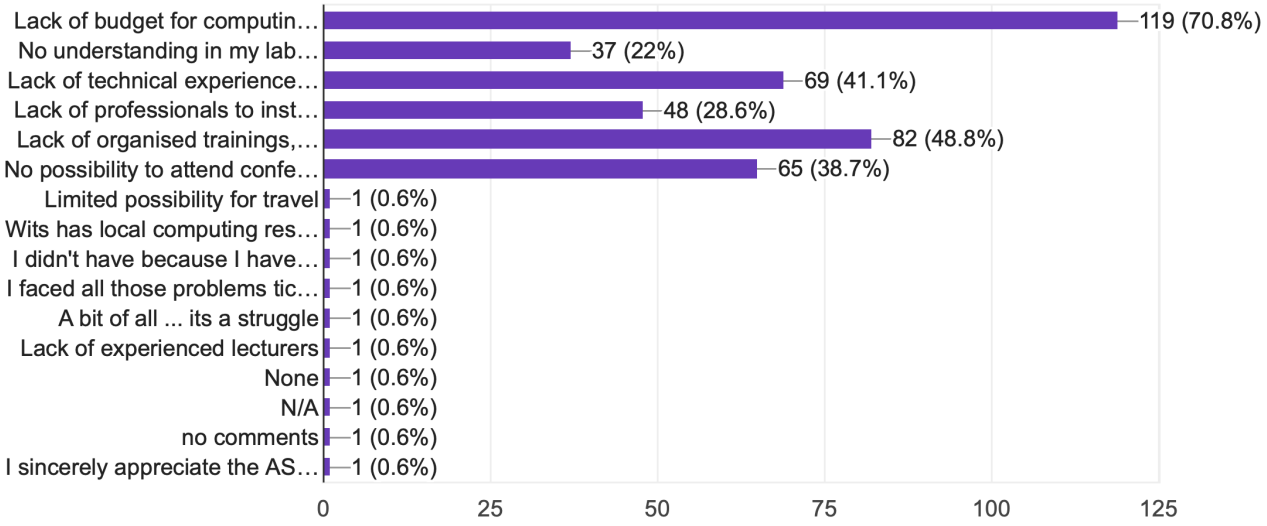


Figure 6 : Problems faced by the scientists to achieve their work

- Computing resources: the largest number of responses stress as well the lack of budget for computing, the lack of technical support and the fact that the hierarchy does not understand the need of computing for research.

- Education: the participants point out the lack of organised training and workshops and the difficulty to attend those organised abroad. More detailed information is found in fig 7 below about the teaching of computing: 74% of the scientists are not provided courses and lectures, or at an insufficient level.

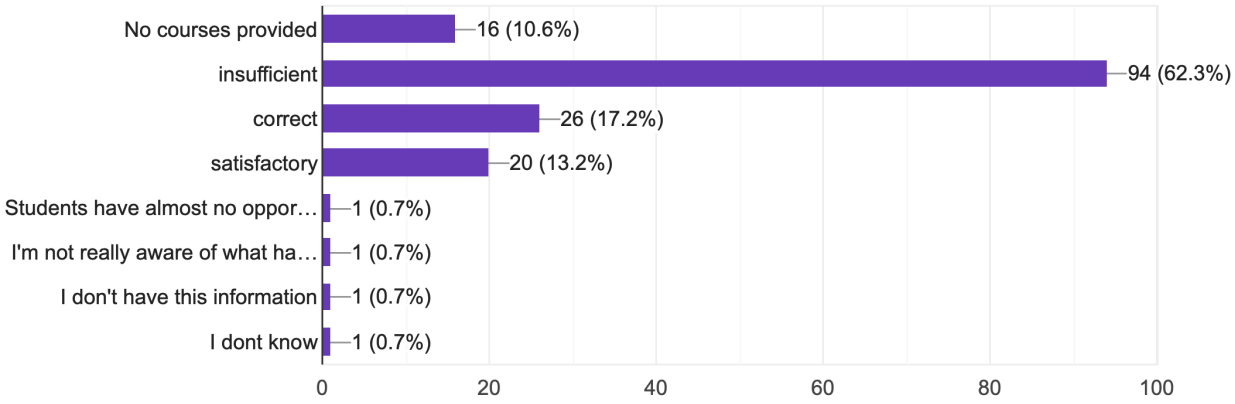


Figure 7: Teaching and training level.

7. General comments

In the free comments gathered for some questions, the participants highlight and give precisions on the points raised above in particular about the resources: sometimes it is 1 computer that is needed and not being provided. HPC resources are alos cited, certainly due to the needs related to AI and DL.
Many others raise the points of lack of budget, lack of professionals to install and run data centres and the difficulty to find collaborators to work in team.

Being asked what they would consider to improve research productivity in their country, the participants elaborated on the fact that African scientists would gain a lot by working together and on the importance of having collaboration within African countries as well as with foreign countries.

8. Conclusion

This survey was launched to evaluate the status of computing resources in the field of African physics research. The panel is mainly composed from participants from Africa and residing in Africa.

- **Infrastructure**: the infrastructure should be improved at a significant level.

- One essential part of the Computing situation is the access, availability and performance of the Network, i.e., Academic and Research Network, in Africa. Networks are vital for the access to data and information. This is not only true at the level of the laboratory but even more at the level of the country and with other countries. We need to have a global picture of the Network status in order to know the possible problems for the research groups. Without filling this gap, there is no possibility of collaboration or share of knowledge.

- Storage and computing power are necessary to store and process the data, which is the only way to produce results and science. The computing needed is more and more sophisticated now that Artificial Intelligence and Deep Learning have entered the game in all sciences. As suggested by some of the participants, large data centres shared within a country or with other countries within Africa would certainly be a solution that would federate the resources, decrease the costs and the disparities between universities and countries.
- Qualified technical staff are necessary to deploy and run these computing resources and make them available to the physics research scientists that would not be able to deal with Cloud deployment or computer access to storage. Here a collaboration between different African countries and foreign countries could be a fruitful initiative to share IT technicians, setup few test sites, and start having an infrastructure on site.

- **Knowledge**: The poll has highlighted the insufficient level of education in computing. Many solutions could be envisaged simultaneously:
  - Increasing the level of computing courses in the cursus of the physics (and other sciences) students.
  - Training IT professionals to prepare the infrastructure.
  - Organising more workshops and training. This would be highly beneficial for knowledge sharing and knowledge update to stay in the forefront in computing where evolution is very fast. But this would have an important positive side effect: Researchers have highlighted the fact that they quite often work isolated. These workshops are the best place to meet their peers and initiate collaborations that would only raise the research productivity.
  - Last but not least, collaboration with others more advanced in these fields throughout the world might speed up the knowledge transfer and build collaborations that would be mutually beneficial.

Raising the awareness of the governing bodies at each level, advisor, universities and state, about the importance of the computing in physics research is absolutely necessary as these changes would certainly need budget and strategic planning.
Budget should be expressly devoted to computing. None of the main discoveries of the last decade would have been possible without the simultaneous use of many powerful data centres throughout the world.