CERN Doctoral Student Programme

A global and distributed data challenge

Many large scale scientific experiments are reaching a breaking point where the growth rate of the collected data greatly exceeds the growth rate of the infrastructure behind them. In the next few years, large instruments similar in scale to the Large Hadron Collider (LHC) and its High-Luminosity upgrade HL-LHC are coming online, such as the Deep Underground Neutrino Experiment (DUNE), or the Square Kilometre Array (SKA) radio observatory. Throughout their lifetimes, these collaborations anticipate massive increases both in the number of data objects they need to handle as well as the total volume of data they need to store. Additionally, increasingly complex computational workflows result in similarly complex dataflows to support them, which can rapidly lead to science-inhibiting complications. Examples include congestion on networks, disorderly space allocations on storage, or erratic transfer schedules. At the same time, there are many smaller communities and experiments who do not want to lose efficient access to the same shared storage and network resources but do not possess a similar level of effort to ensure their sustainability. This brings a variety of challenges to the field of distributed data management as a whole to ensure fair use of all available resources, leading to the need to orchestrate and synchronize data flows across experiments with competing characteristics.

Become a doctoral student at CERN

The Computing Section in the ATLAS Data Processing group of the CERN Experimental Physics Department is actively and continuously searching for bright and motivated computer science students who want to undertake a funded doctoral study up to 3 years in the area of distributed data management. The topic of the study will be determined together by the candidate, CERN, and the supporting university. The academic process must be undertaken under the rules and regulations of the university, and will be co-supervised by a professor at the university and a staff member at CERN. Experience with Linux, databases, and common open-source technologies is necessary. Experience with the scientific process, such as publishing of articles, is advantageous.

Get in touch with us

Please direct your inquiries at any time to Mario.Lassnig@cern.ch

Further reading

Barisits, Beermann, Berghaus, et al., *Rucio: Scientific Data Management*, Comput Softw Big Sci 3:11, Springer, 2019, https://doi.org/10.1007/s41781-019-0026-3

Lassnig, Barisits, Christidis, *Experiences in Exascale Scientific Data Management*, IEEE Computer Society, Data Engineering Bulletin 43:1, 2020 http://sites.computer.org/debull/A20mar/p9.pdf