

Preparing CERN Tier-0 data centres for LHC Run 3

*Dr. Olof Barring*¹

¹CERN, 1211 Geneva 23, Switzerland

Abstract. Since 2013 CERN's local data centre combined with a co-location infrastructure at the Wigner data centre in Budapest [1] have been hosting the compute and storage capacity for Worldwide LHC Computing Grid (WLCG) [2] Tier-0. In this paper we will describe how we try to optimize and improve the operation of our local data centre to meet the anticipated increment of the physics compute and storage requirements for Run 3, taking into account two important changes on the way: the end of the co-location contract with Wigner in 2019 and the loan of 2 out of 6 prefabricated compute modules being commissioned by the LHCb experiment [3] for their online computing farm.

1 Background

In early 2019, CERN's computing facilities were comprised of a 3.8 MW central Data Centre facility and co-located IT capacity in the Wigner Data Centre in Budapest (HU). The combined facilities hosted about 1 000 racks with 15 000 servers and 230 000 processing cores, more than 350 000 terabytes of raw disk-based storage capacity on 100 000 disk drives, and some 350 000 terabytes of tape-based storage capacity.

The co-location contract with Wigner Data Centre was ending with 2019 and the original plan was to compensate with a new computer centre at the French (Prévessin) site of CERN. A call for tender for a Turn-key Data Centre was issued in 2017 but unfortunately subsequently cancelled before the contract was awarded. Consequently CERN IT department had to quickly find alternatives to compensate for the imminent loss of Wigner co-location capacity as well as the additional resource requirements for the next LHC run (Run 3), which is scheduled to start in 2021.

In parallel the ALICE and LHCb experiments at the LHC were upgrading their trigger farms for Run 3. They conducted a joint tender for pre-fabricated data centres. The winning bid was based on container-like modules, each with 24×48U* racks and 500 kW of IT equipment load. The solution provided a power usage effectiveness (PUE) [4] of less than 1.1 thanks to indirect free air-cooling with adiabatic support. LHCb acquired 6 modules but since they would only need 4 of them for Run 3 they generously offered us the use of the remaining two modules.

* U (rack unit) is a rack height measure. 1U = 44.45 mm (1.75 inches)

In the rest of this paper we will describe how we gradually reduced our capacity at Wigner and repatriated it to CERN for recommissioning in both our current main data centre and the two LHCb compute modules.



Figure 1: Schematic floor plan of R-050. The rack colours reflect the occupancy (green is empty).

2 Central computing facilities

The central computing facilities at CERN are hosted in a single building on the Meyrin site. It was built in the 1970s for hosting mainframe computers. After a number of refurbishments throughout the years it currently consists of three separate IT equipment rooms with slightly different features:

- R-050: 1445 m² floor-space, 659 racks and 2.1 MW power (see Figure 1);
- R-060: 228 m² floor-space, 90 racks and 450 kW power;
- S-034: 1241 m² floor-space, 260 racks and 1.1 MW power;

The two large rooms (R-050 and S-034) are mainly hosting the “physics processing” of the different experiments. The small room, R-060, is hosting critical services of standard business computing type (administrative systems and databases, email, etc.).

All power feeds are UPS backed up with an approximate autonomy of 10-15 minutes. In addition, for the critical services hosted the R-060 room has two independent power feeds, of which one also has diesel generator backup. The majority of racks are nowadays equipped with intelligent (switched and metered) power distribution units (PDUs), which enables a fine-grained monitoring of power usage.

The R-050 and S-034 rooms are air-cooled with rack row pairs arranged in aisles for cold-air containments. Each aisle consists of 20-30 racks. The cold air is produced in chillers at the roof of the building. The cold air is blown through large ducts along the walls

to under the false floor where it diffuses into the cold aisles through perforated floor tiles. The average power density supported is approximately 4 kW per rack.

In the S-034 room there are also 80 racks with water-cooled heat exchangers in the rear door, which allows for a power density of 10 kW per rack. The same type of racks are used in the R-060 room.

3 LHCb compute modules

Each compute module has 24 48U racks and supports a total power of 500 kW, which results in an average rack power density of ~21 kW for a fully loaded module. The electrical feed is a single 66 kW source feeding two transformers providing dual independent feeds to each rack. There are no UPS systems or diesel generator backup on any of the feeds, which means that even a short power cut would normally result in an immediate non-graceful stop of the IT equipment.

The modules were gradually deployed and commissioned starting in early 2019 and we got access to the first of our two modules in March.

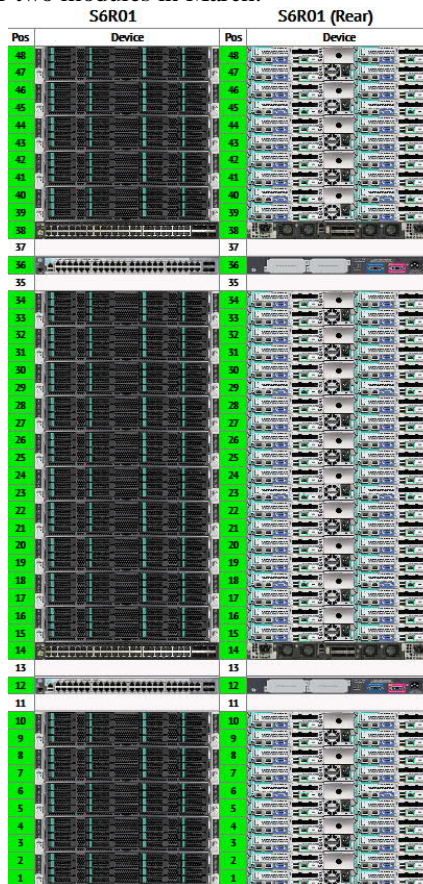


Figure 2: typical rack layout in the compute modules.

Because of the lack of UPS we decided to not deploy any storage arrays. The relatively high power density also favoured compute intensive batch workload. The standard rack deployment, Figure 2, consisted of 20x 2U chassis with four dual-socket servers each

together with two data switches and two management (IPMI) switches, which filled the 48U available. One rack out of the 24 is used for the network routers.

4 Repatriation of IT equipment from Wigner

CERN IT department had been co-locating part of its compute and storage capacity at the Wigner Data Centre (WDC) in Budapest since 2013 [5]. By the end of 2018 we had about 1000 2U chassis with 4 servers each (2U4N) and 900 disk arrays (JBODs) deployed at WDC. In order to prepare for the end of the co-location contract at end of 2019 we decided to repatriate the majority of the equipment in 6 shipments throughout 2019, see Table 1.

Table 1: Repatriation of IT equipment from Wigner to CERN in 2019.

Shipment	2U4N	JBODs	EU Pallets
February	150	0	15
April	54	72	19
June	129	144	33
July	150	104	29
October	174	120	34
December	183	0	20
Total	840	440	150

Out of the 3360 (840×4) servers, 220 (55 2U4N) were used as front ends to the JBOD storage arrays that were installed in our central computing facilities. The remaining 3140 servers (785 2U4N) were installed in the LHCb compute modules. The decommissioning of equipment is clearly visible in the power consumption for our co-located IT equipment at Wigner, see Figure 3.

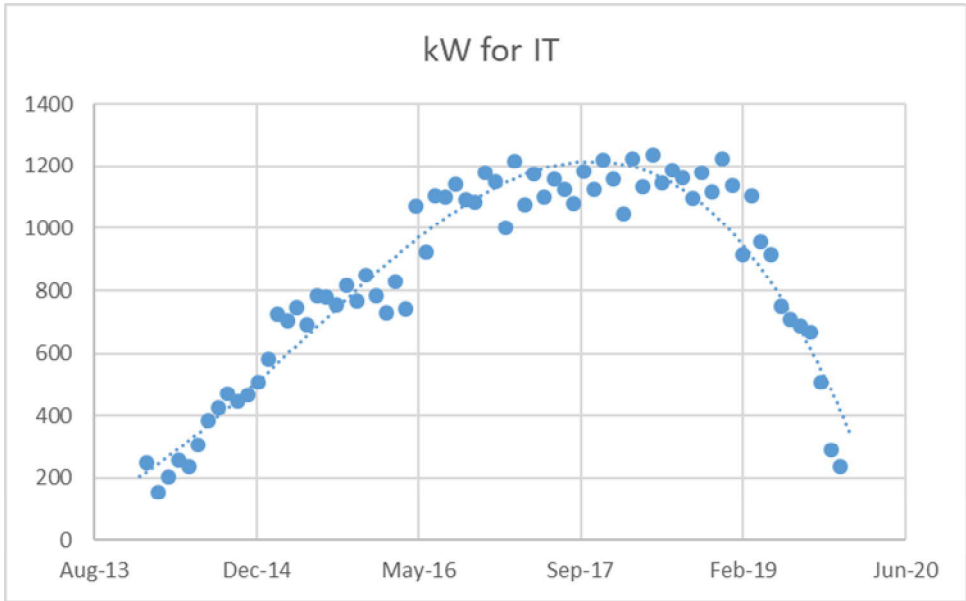


Figure 3 : Power consumption (kW) for our co-located IT equipment over the years.

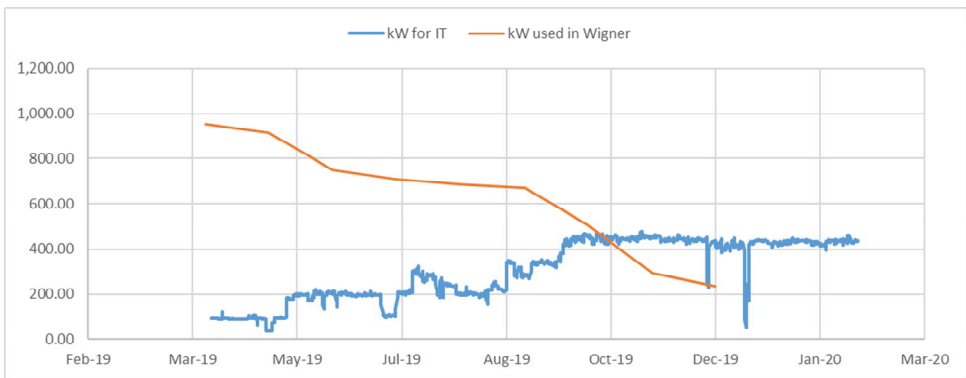


Figure 4: Power consumption (kW) in first LHCb compute module (blue line) and in co-location facility (orange line).

The corresponding ramp-up of the first compute module at LHCb is shown in Figure 4, which for comparison also includes the tail of the power consumption of the co-located IT equipment in Wigner shown in Figure 3. In the figure one can also see two power cuts experienced in late November and mid-December. The first one was a general power outage at CERN caused while the second was a scheduled maintenance.

5 Experience with LHCb compute modules

Thanks to LHCb collaboration's kind offer to let us use two of their compute modules we managed to maintain a relatively stable compute capacity for the experiments through 2019. Evidently there were periods with slightly reduced capacity while the equipment was drained, dismantled at Wigner and shipped to CERN but the impact was minimal thanks to very good support from Wigner. For every shipment we decided to run our standard burn-in

tests of the equipment to assure it had not been damaged during the transport but general there were none or very little (a few JBOD disks drives).

The compute modules have functioned very well with the dense deployment described above (Figure 2) and it can be seen in Figure 4 that since the module was completely filled during the autumn 2019 we have been running at between 80-90% of the 500 kW available.

One initial concern was that the lack of UPS may increase the risk for damage of the equipment in case of a power cut but this turned out to not be the case after the unscheduled cut in late November. It should be noted that we only deploy servers and most of the servers are nowadays equipped with solid state drives instead of mechanical (spinning) disks. We would likely have experienced more damage with the JBODs.

One disadvantage with the compute modules is that becomes very difficult work inside them, e.g. mount servers into the racks, once they are already commissioned with live equipment. Apart from the noise it is a windy environment due to the high air flow generated by the four large Air Handling Units (AHU) at the roof. The temperature at the rear of the racks is also quite high (~40-45°C), which makes the cabling of a neighbouring rack quite inconvenient.

6 Outlook for Run 4

In 2019 the CERN Finance Committee and subsequently also the Council approved the funding for a new Data Centre at CERN. The Market Survey has been launched and the invitation to tender is planned to go out in Q1 2020, which would allow for the contract award by end of 2020 and commissioning in 2023. We aim for a turn-key data centre with an initial 4 MW IT equipment load upgradeable in stages to a maximum of 12 MW. The Power Usage Effectiveness (PUE) is required to be ≤ 1.1 , which is facilitated by allowing for inlet temperature excursions up to 32°C (2011 ASHRAE TC9.9 A1 [6] temperature environment). The average rack power density shall be 20 kW but with limited (10-20%) UPS coverage.

7 References

- [1] “Wigner Datacentre”, <https://wignerdc.wigner.hu/home>.
- [2] “The LHCb collaboration”, <https://lhcb.web.cern.ch/>.
- [3] “Worldwide LHC Computing Grid”, <https://wlcg.web.cern.ch/>.
- [4] ISO/IEC, “Power usage effectiveness (PUE)”, ISO/IEC 30134-2:2016(E), 2016.
- [5] O. Barring, E. Bonfillou , B. Clement , M. Coelho Dos Santos, V. Dore , A. Gentit, A. Grossir, W. Salter, L. Valsan and A. Xafi, “Experience with procuring, deploying and maintaining hardware at remote co-location centre”, *Journal of Physics: Conference Series*, vol. 513, no. 6, May 2014.
- [6] ASHRAE: American Society of Heating, Refrigerating and Air-Conditioning Engineers, ASHRAE TC9.9, Thermal Guidelines for Data Processing Environments, fourth ed., 2015, p. 164.