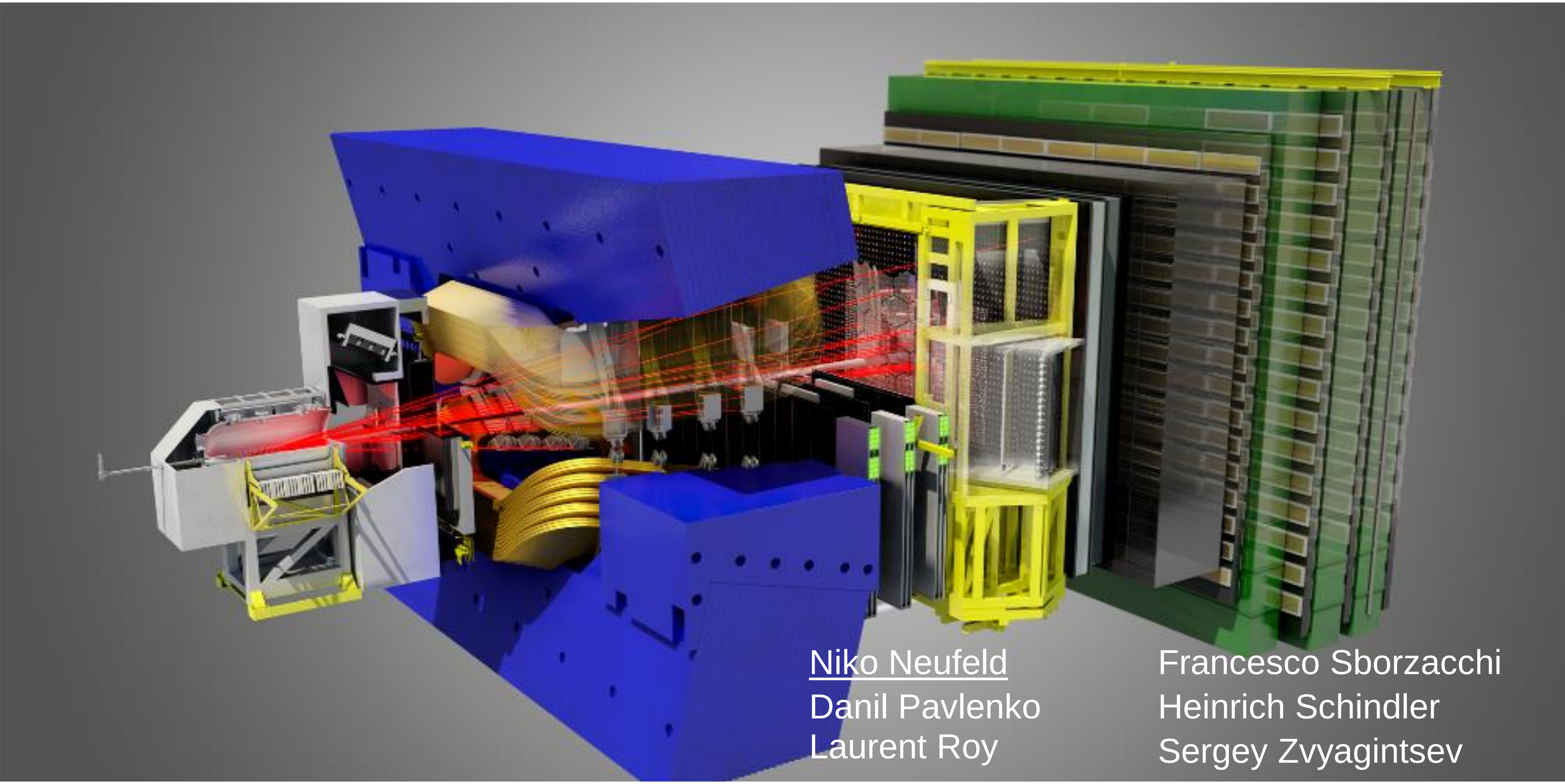


Niko Neufeld
Danil Pavlenko
Laurent Roy
Francesco Sborzacchi
Heinrich Schindler
Sergey Zvyagintsev

CERN, EP Department

Sustainable computing solutions: a case-study of the LHCb data-centre

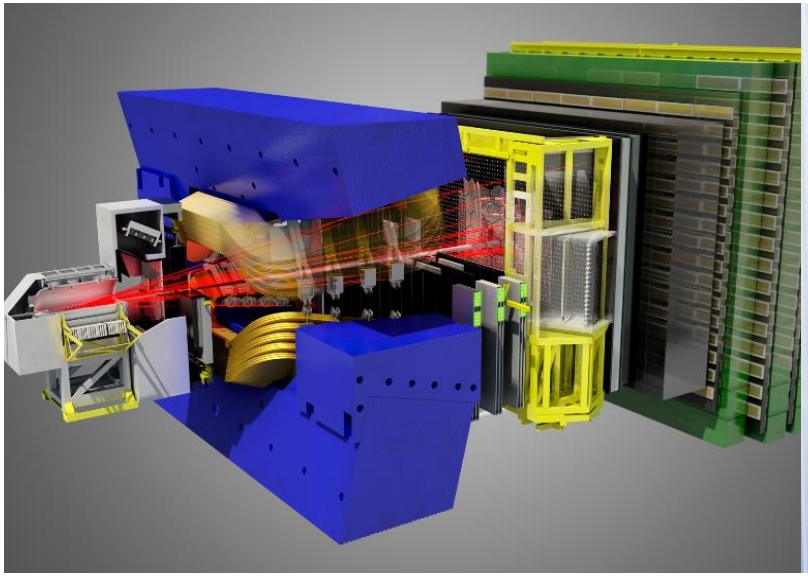
ICHEP 2024
Prague

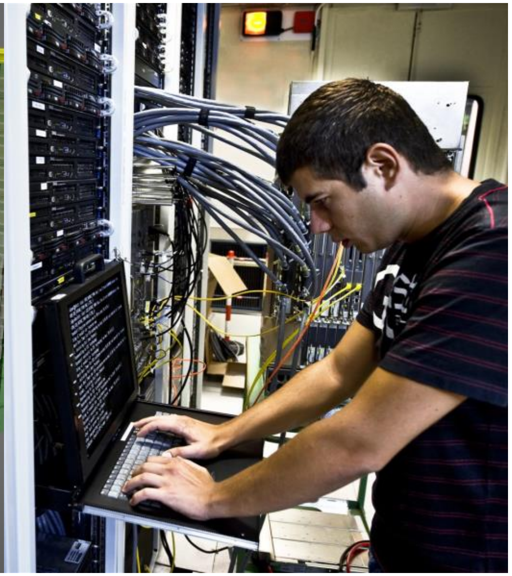
Acknowledgements & Disclaimers

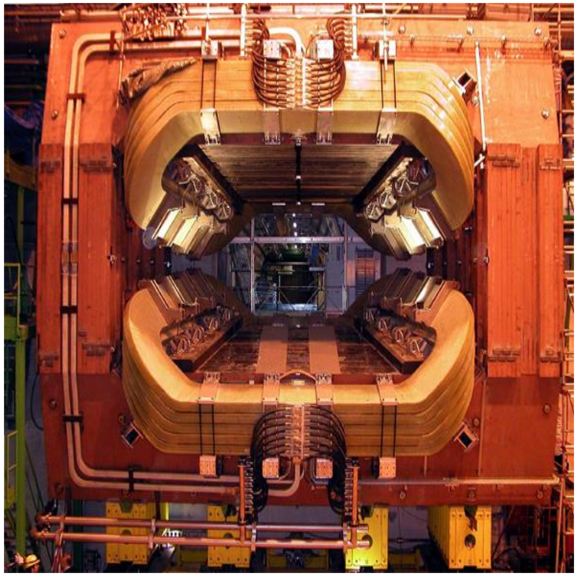
Much of the material presented has been prepared by my colleagues Laurent Roy (CERN) and Francesco Sborzacchi (CERN). I'd like to thank them

Power-saving is only a part of LHCb's comprehensive environmental impact strategy

Why do we care about computing?





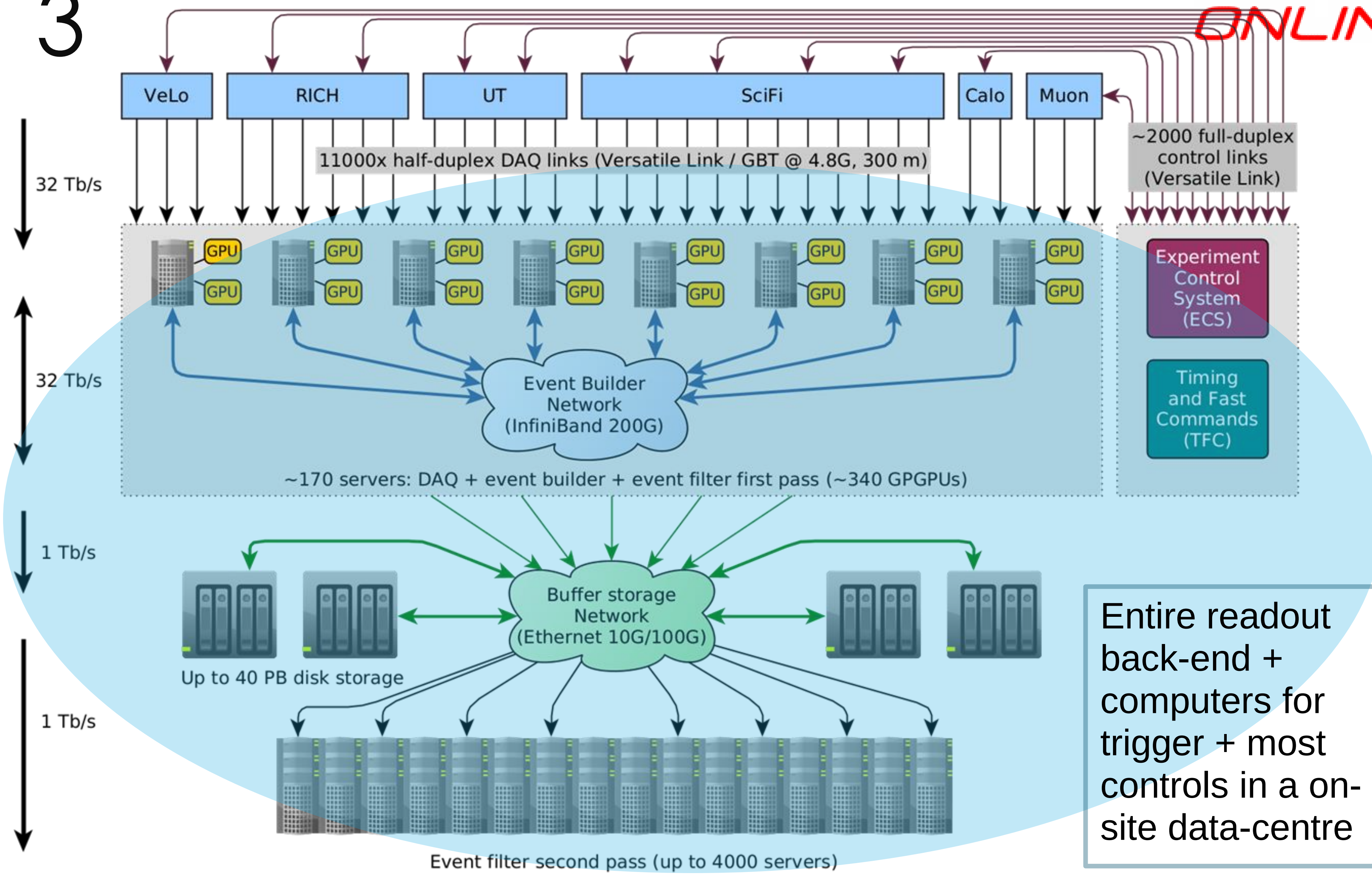


LHCb detector < **0.3 MW** (excluding the magnet)

LHCb DAQ + High Level Trigger **1.4 MW**

LHCb Magnet **4.4 MW**

LHCb DAQ and High Level Trigger in LHC Run 3



Entire readout back-end + computers for trigger + most controls in a on-site data-centre

Side remark – the value of density

- Physical proximity is one of the most effective means to safe cost in data-centre
- Fast connections (> 10 Gbit/s) are usually factors cheaper when they remain within 3 meters. *Short connections also consume less energy*
- It is beneficial to have a high density of equipment, this drives the power-dissipation density up
- There is a completely analogous trend **within** the compute equipment itself: it is much more power- and cost efficient to have more cores, logical elements, memory in a single CPU, GPU, FPGA package —> consequently modern high-end chips dissipate 500 Ws and more(!)

Energy usage in a data-centre

- The data-centre infrastructure: cables, fans,, lights, pumps, water-treatment, electrical switch-boards
- The computers: CPUs, power-supplies, fans, memories
- Add-in cards: GPUs, NICs, FGPAs,
- Storage: drives, controllers,
- Network equipment: switches, physical interfaces (PHYs), cables

Core metric = Power Usage Efficiency

PUE =

$$\frac{\text{Total } power\ consumed\ by\ the\ data{-}centre}{\text{Total } power\ consumed\ by\ IT\ equipment}$$

How can we do better?

- Less compute
- More energy-efficient compute
- Energy re-use
- Energy efficient data-centre / infrastructure —> this talk

How to cool?

Main heat dissipating unit are compute servers pushing hot air out at the rear of a metallic enclosure

Network equipment behaves the same for the purpose of this consideration.

We have disregarded direct liquid cooling solutions for the time being since it seemed to early at the time of planning (2017)

==> Head-load is all hot air :-) which needs to be cooled

Can be done directly (expansion cooling, quite inefficient) or by taking the air into water or by evacuating the hot air into the environment (“free cooling”)

- **Water** cooled passive or active (with fans) **doors** mounted on the rear of racks
- Necessary piping work at LHC Point 8 made these solutions cost-wise unattractive

Server needs

Defined by [ASHRAE](#) (American Society of Heating, Refrigerating and Air Conditioning Engineers) recommends a maximum of 27 °C (low consumption, risk of failure very low). But have also defined environmental envelopes (Class A1, A2, A3, A4)

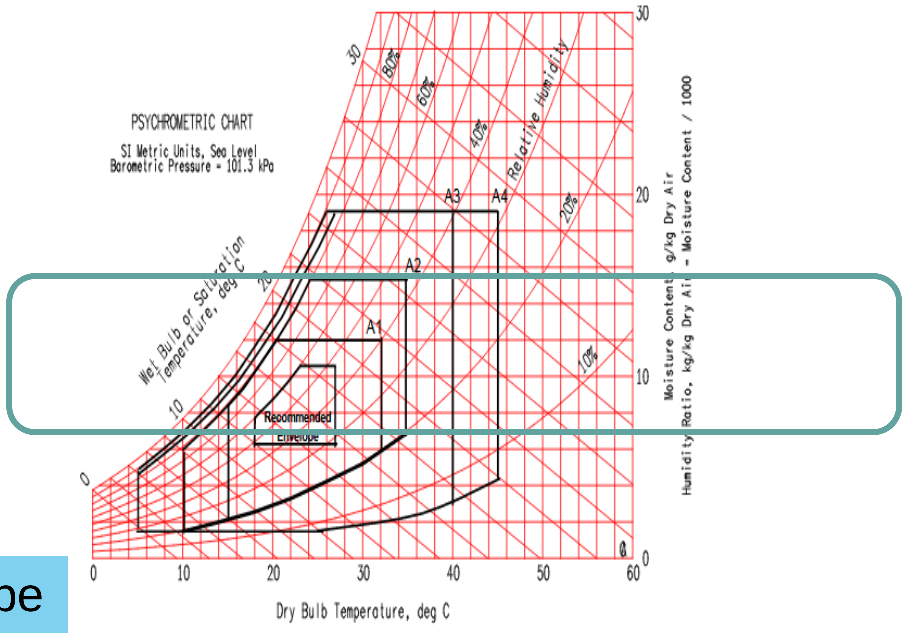
The manufactures use the Class to define their specs.

Servers compliant with 'ASHRAE Class A3 or A4' are available, but less common

Also servers dissipate more heat when running at higher inlet temperatures

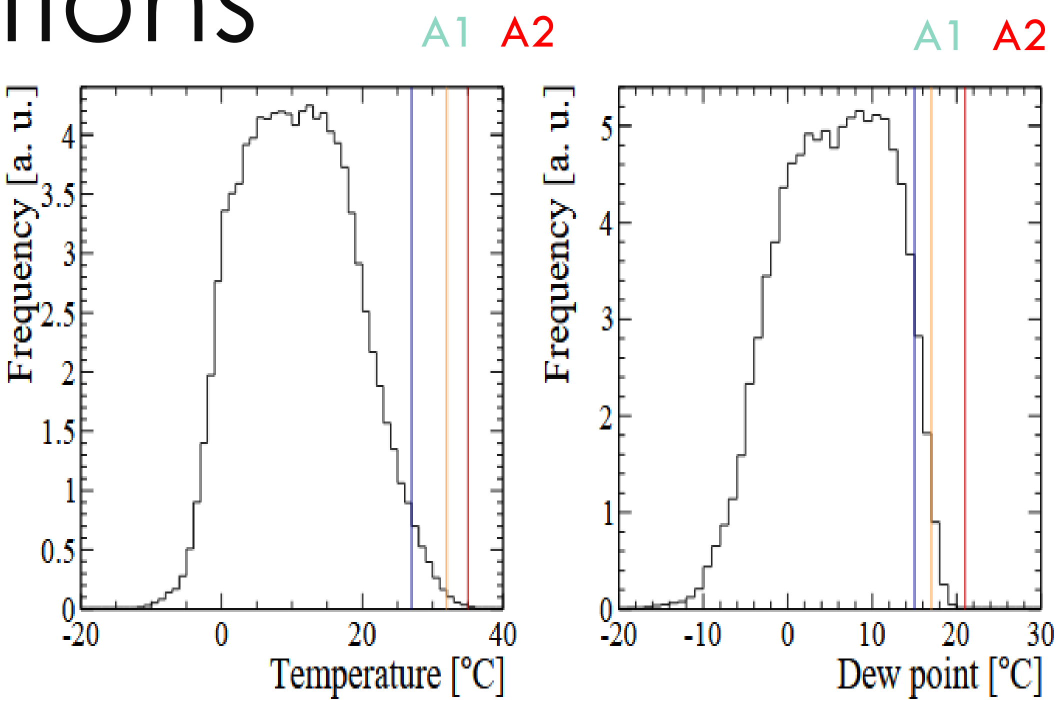
LHCb targets PUE < 1.1 within A2 envelope

	Temperature range [°C]		Humidity range		Max. dew point [°C]
Recommended	18	27	5.5°C DP	60% RH, 15°C DP	
A1	15	32	20% RH	80% RH	17
A2	10	35	20% RH	80% RH	21
A3	5	40	-12°C DP, 8% RH	85% RH	24
A4	5	45	-12°C DP, 8% RH	90% RH	24



Geneva outside AIR conditions

Distribution for 1 year



- (Almost) Always in the A2 envelope
- Compatible with 'Direct Free AIR cooling'
- Compatible with 'Indirect Free AIR cooling' (with additional adiabatic cooling for summer)

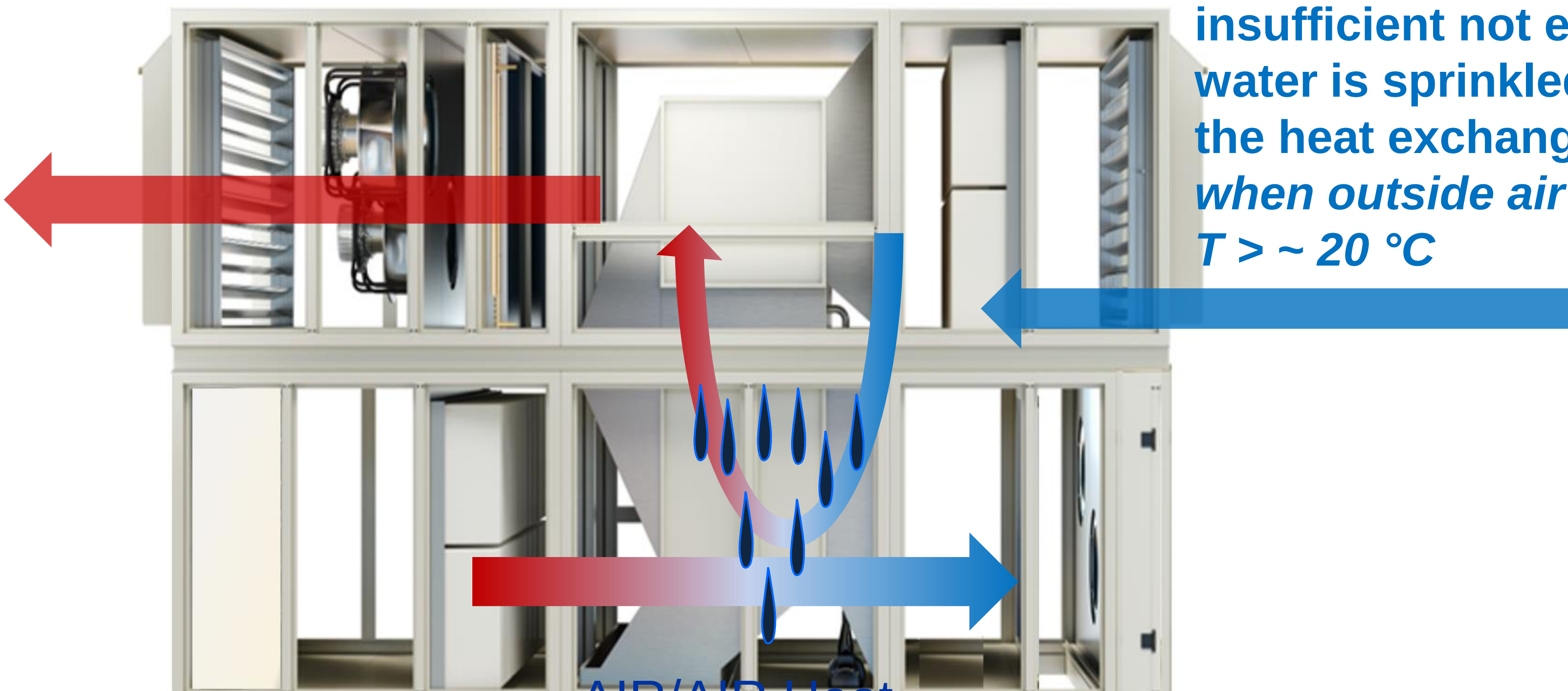
  advantages with 'Indirect' solution: Air filtration easier, better control of the air temperature inlet.

Indirect Free Air Cooling

LHCb ONLINE

Outside Fans running 0-20% - 70%

If outside airflow insufficient not enough, water is sprinkled on the heat exchangers *when outside air T > ~ 20 °C*



Return Air

AIR/AIR Heat exchanger

Supply Air

> 50°C

Inside loop: fans running according to DC Load

27°C

Winning solution after tendering

Modular Data Centre’ consisting of single row modules (shaped like a container but larger than standard shipping container)

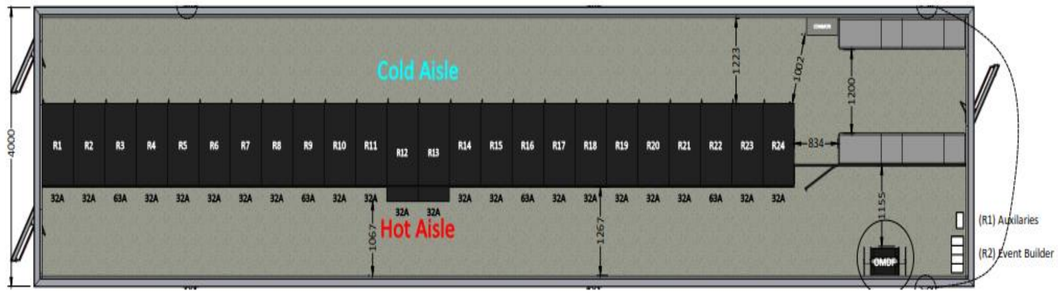
LHCb: 7 Modules
6 IT: 18m x 4m x 6,5m height
1 Power + Water : 18m x 3,5m x 3m height

- Up to ~ 20kW/rack (85%)
- Up to ~ 40kW/rack (15%)
- Designed for a total power of **2.3 MW**

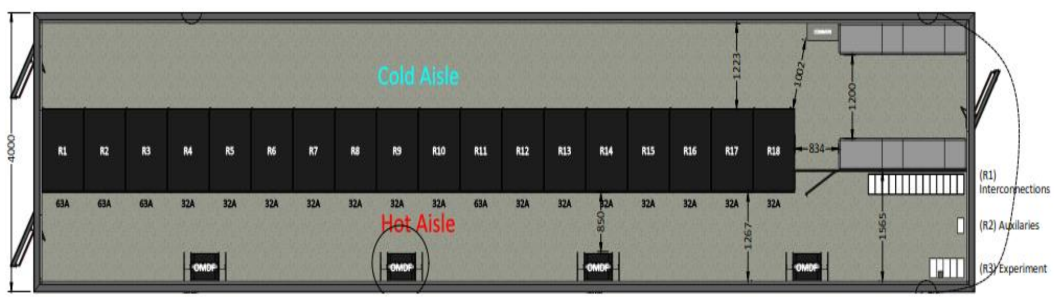
2 full redundant Power supplies (2 transformers 3.15 MVA, 2 main switchboards, 2 secondary switchboards, 2 PDU per rack)

Layout

- Two event-builder modules with 18 racks (800 mm wide).
- Four event-filter modules with 24 racks (600 mm wide).
- 132 racks in total (6336 rack units).



Event-filter module.



Event-builder module.

Transformers & Power module

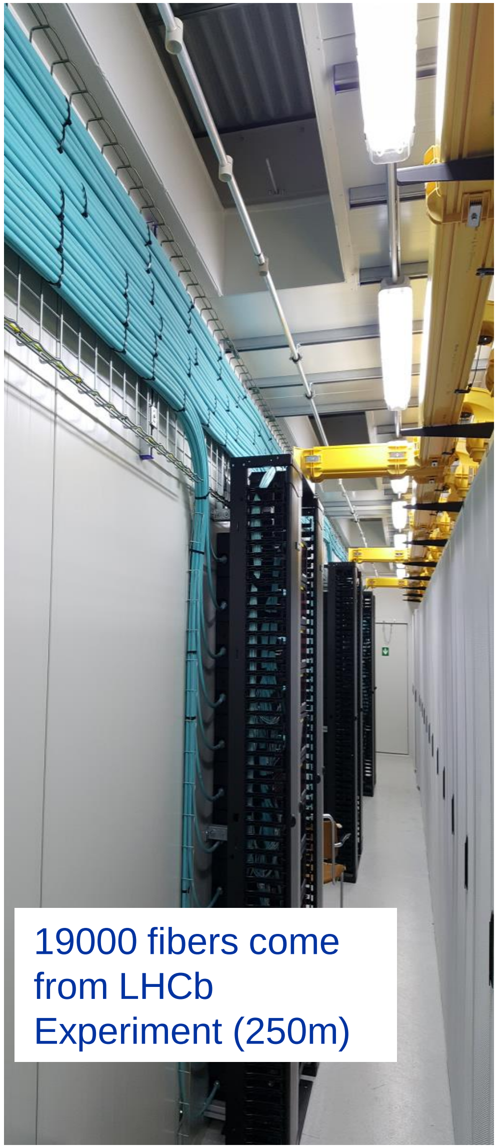
LHCb ONLINE

LHCb site

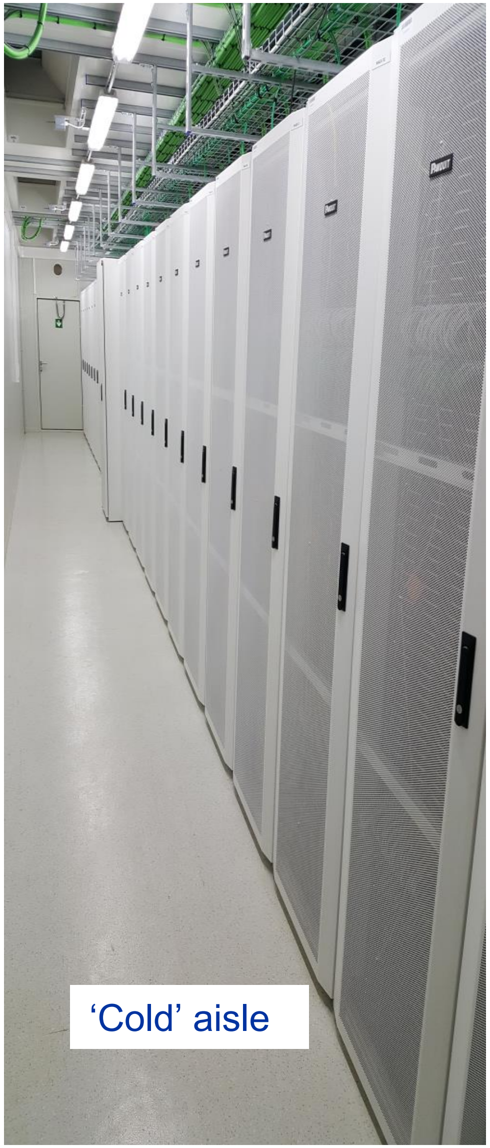
Air Handler Units

Inside the IT modules

Air Unit Controller



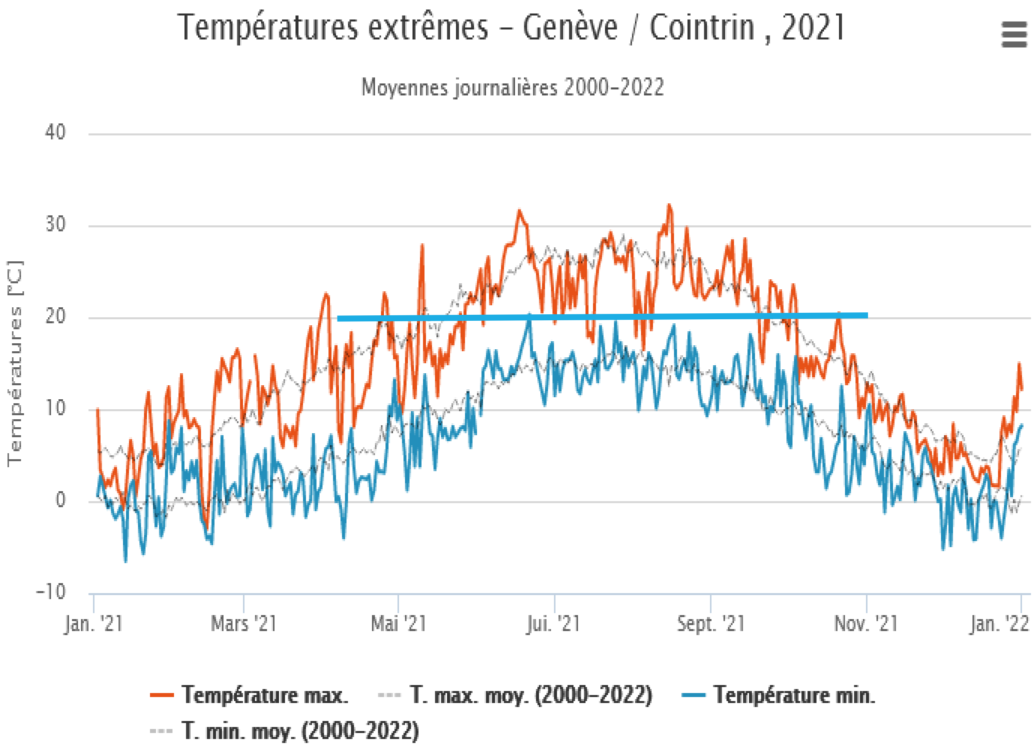
19000 fibers come from LHCb Experiment (250m)



'Cold' aisle

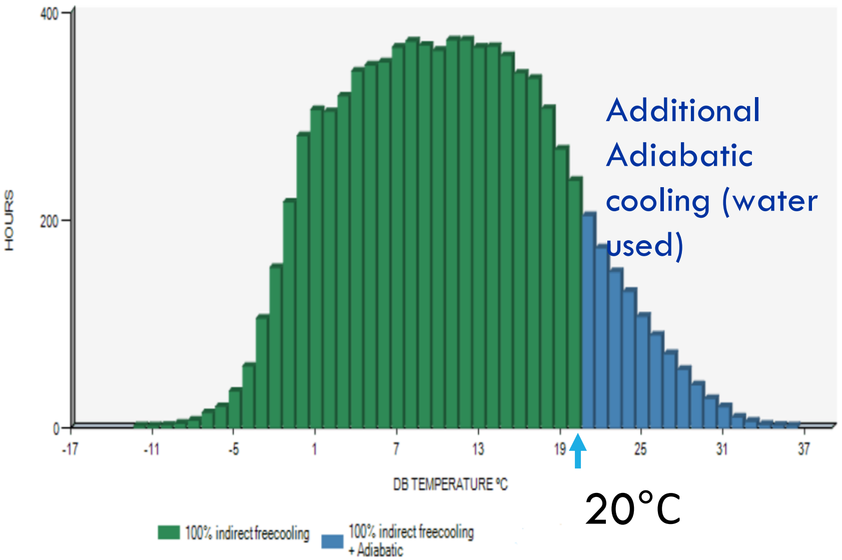
2 secondary switchboards
Smoke detection

Temperatures in Geneva



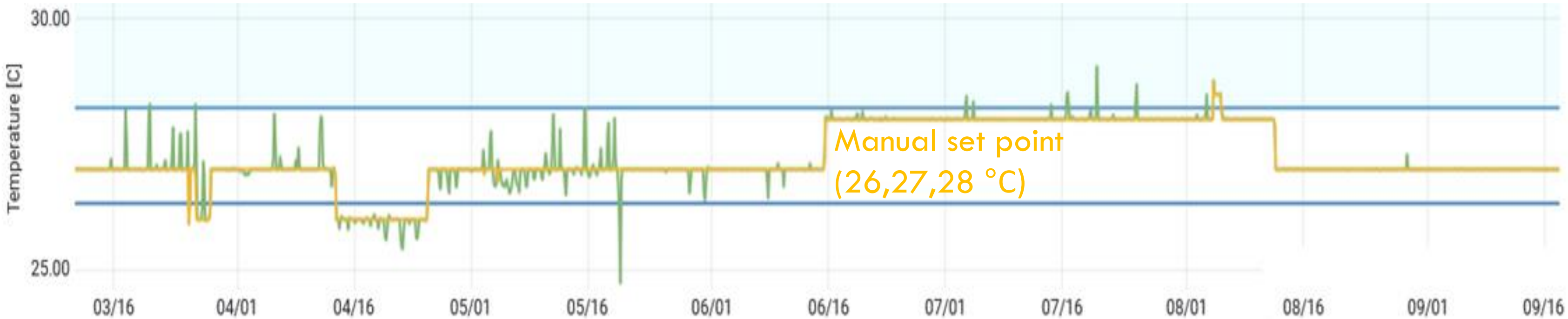
For a full year, the outside air is below 20 °C most of the time

□ 'Dry mode' (without water) is sufficient for a large fraction of the time during a year —> reduces water consumption

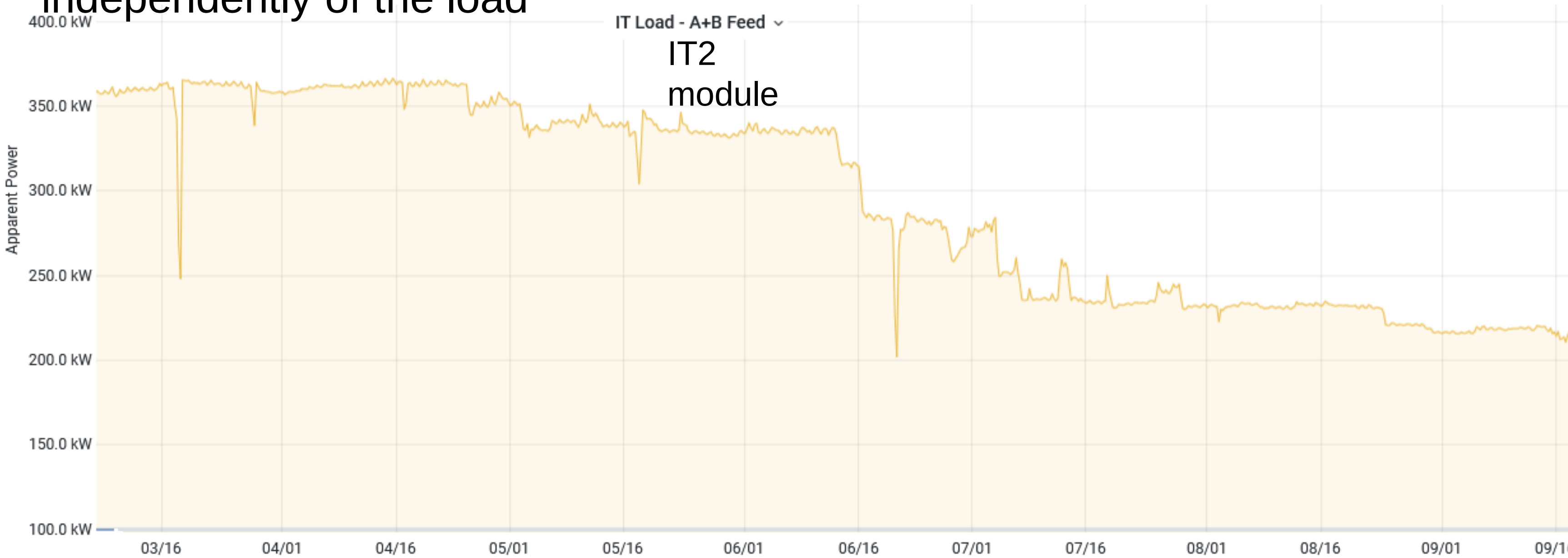


Internal temperature stability

Ex: air supply in IT2 module in 2022



Temperatures relatively stables (+/-1°C) , independently of the load



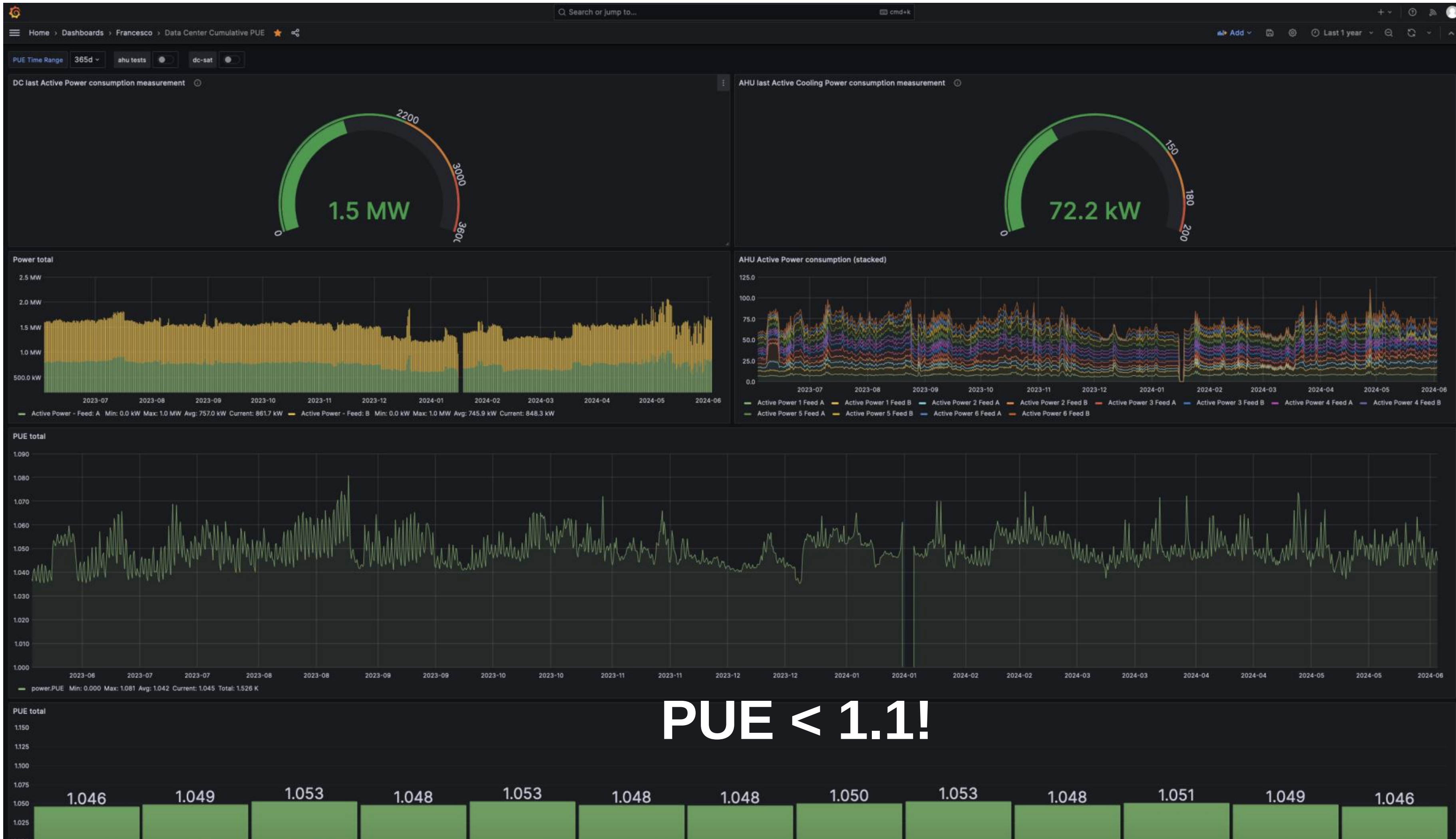
Regulation

Goal □ have a 27 °C inlet server temperature, as stable as possible, independently of **Outside temperature**, humidity & **Load of servers**

- Parameters for each AHU:
- Outside fan speeds
  - minimum 20%, high speeds clog more quickly the air filters
- Inside fan speeds
  - adjusted to have a good delta T temperature for the best Air/Air exchanger efficiency
- Water pumps speeds (in Summer mode)
  - enough to well cover the heat exchanger but not too high (water consumption)
- Influence of the air filters status
  - regular maintenance needed
- Trade-off between PUE and water-consumption

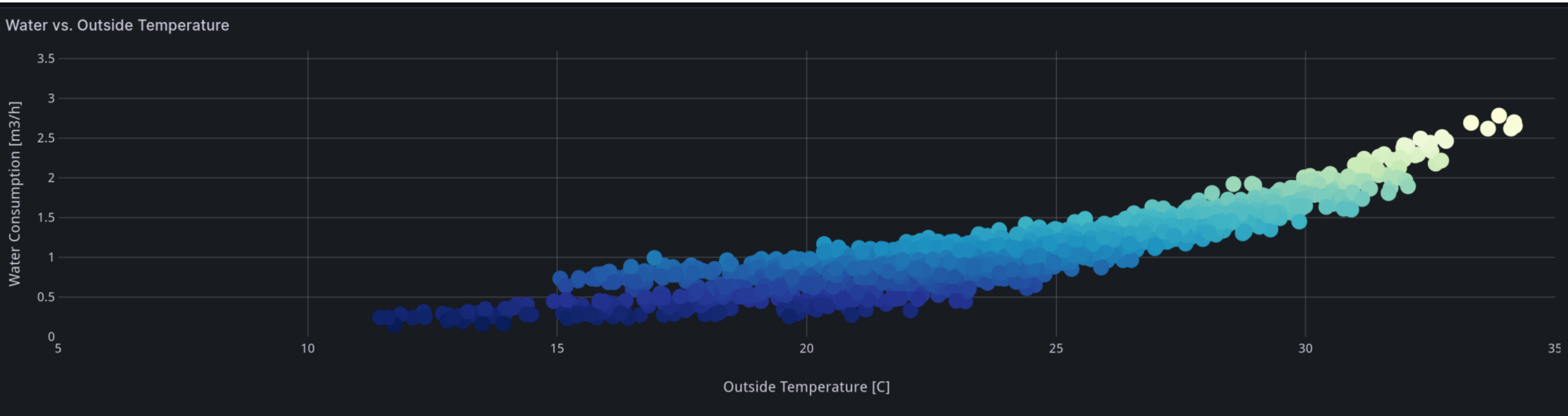
Air filters improvement

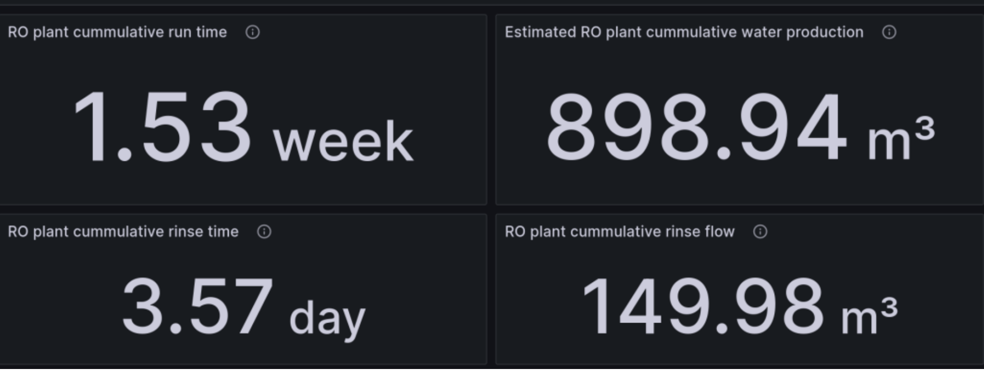
Overall efficiency LHCb DC



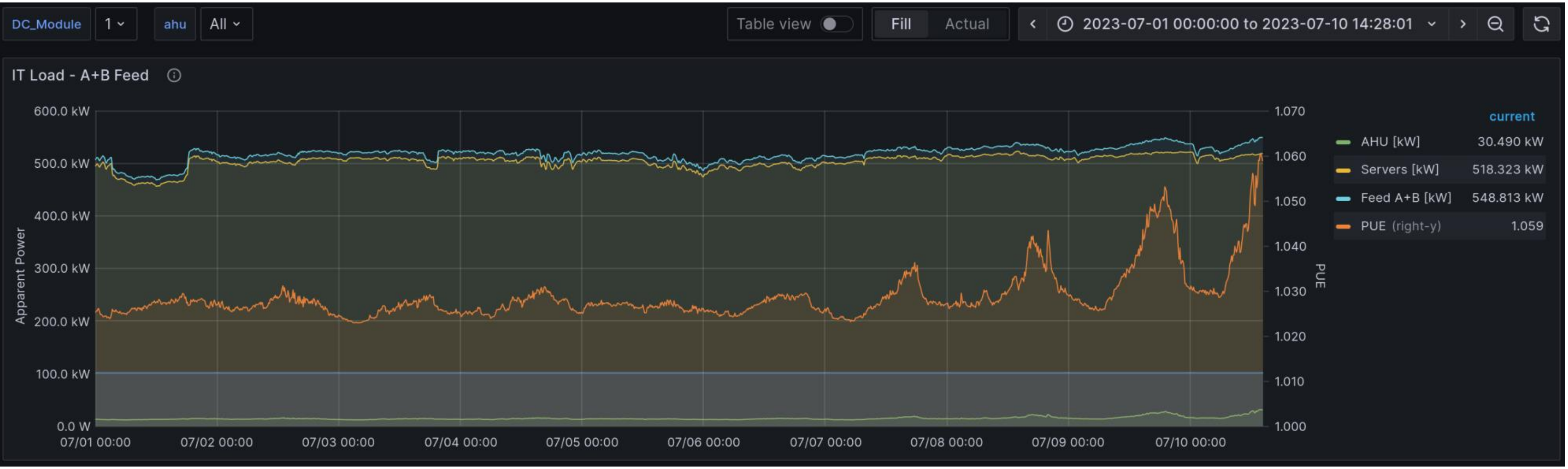
Water consumption

30 days in July 2023
27C set-point interior
1.9 MW average total power





513 kW total load @ 27C vs PUE



Final thoughts

- A lot of factors contribute to the environmental impact of the LHCb Online system
- We have decided to go for a quite efficient infrastructure based on indirect free air-cooling. We consistently achieve a PUE < 1.1
- For the future we will
  - try to reduce water-usage
  - optimise configuration of cluster
  - investigate different CPU / GPU architectures
  - prepare for the use of direct liquid cooling
  - check energy-reuse possibilities with CERN EN department
- Lessons learned will influence final choices for LHCb upgrade II for Run5 and Run6

More material

Energy efficient compute

- We can configure *power-savings options on existing hardware.* Clock-frequencies will be capped and parts of the hardware will be put automatically into a low power-state when they're not used
- Consequences:
  - a certain loss in overall performance (to be measured),
  - a certain loss in "responsiveness" of the system to load-changes
    Most likely irrelevant for asynchronous, batch-style processing
    unclear for quasi-realtime / I/O work-loads
    unacceptable for ECS / slow-controls

Shift to more energy efficient compute

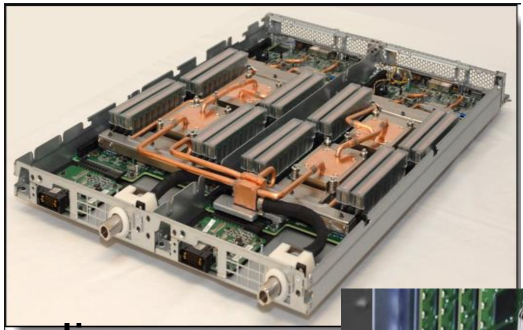
- *Use more energy efficient CPUs* → accept increased price. Example from upcoming Intel server CPUs: for a *44% increase in cost, 25% reduction in energy* consumption (TDP). These numbers are based on nominal power-dissipation and list-prices. *The power savings must be measured with a realistic application and the cost-increase must be probed in a competitive procurement exercise*

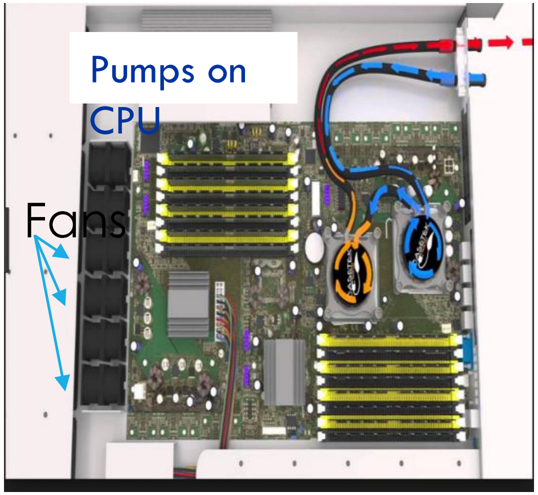
- *Use allegedly more efficient ISAs, ARM, RISCV* —> TCO in terms of money and energy must be determined by benchmarking (not on paper).

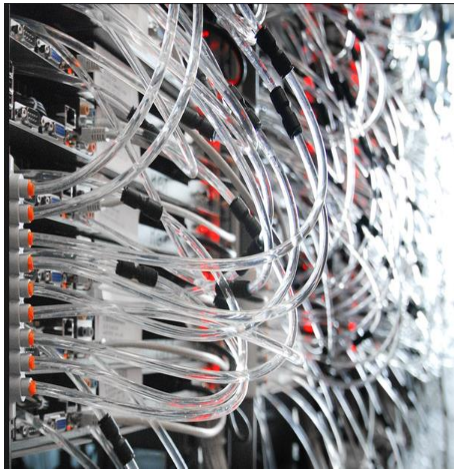
- *Use GPGPUs or FPGAs also for HLT2* → algorithms must be available for a comparison. In GPU-friendly applications the power-savings are typically at least 50%, cost-savings depend again on competitiveness of procurement procedure → for GPGPUs would be crucial to be able to run on Nvidia, AMD AND Intel GPUs. Ceteris paribus same goes for FPGAs

Improved cooling of the servers

- Direct Contact Liquid Cooling
  - Better potential for energy reuse (warm water)
  - Energy savings due to reduced fan-speed (to be bench-marked)







- Immersion Cooling
  - drastic change of infrastructure
  - only rack-level
  - rather higher up-front cost