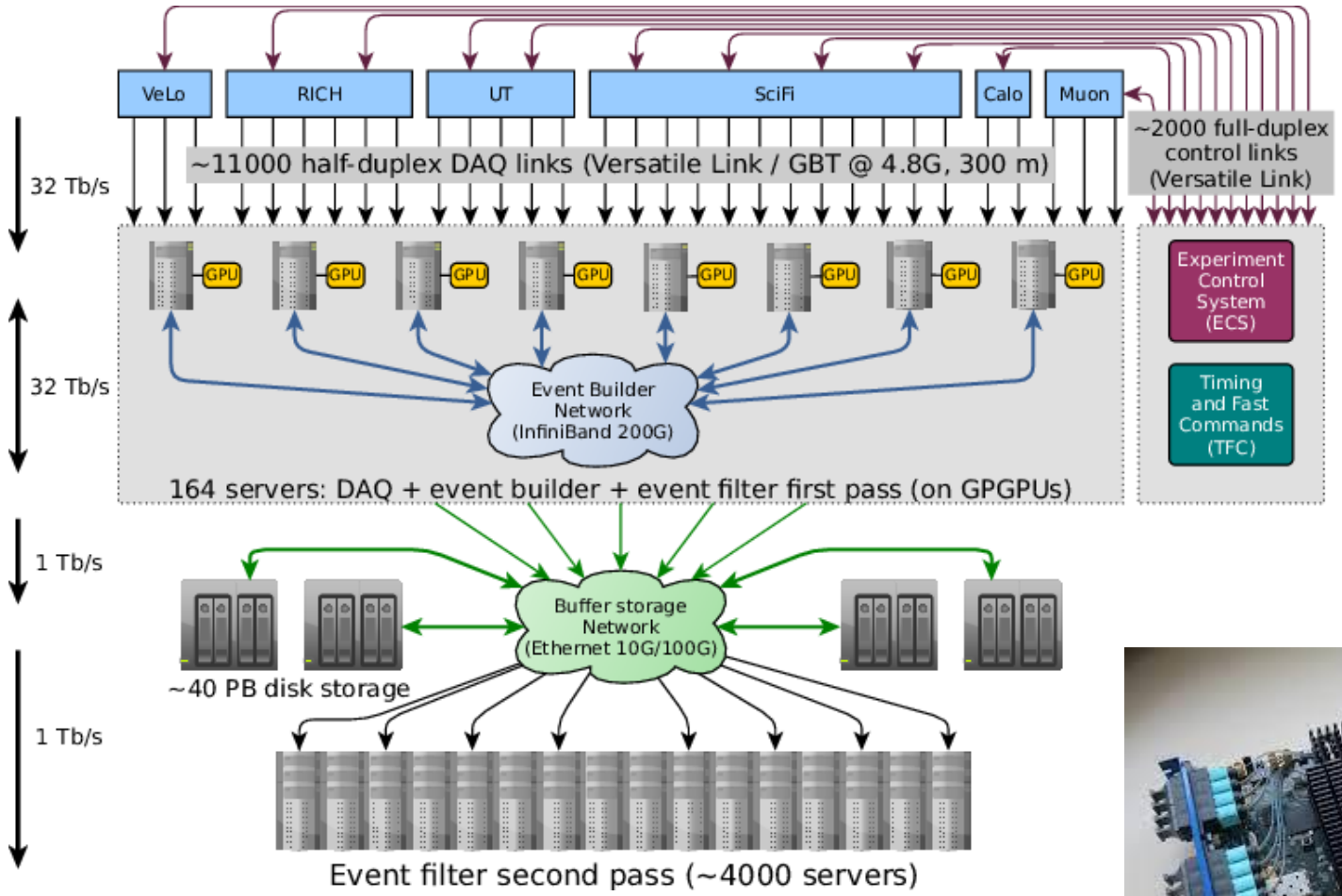
Power Consumption Optimization for LHCb Online and DAQ System

Matías Leandro Flores
Supervisor: Francesco Sborzacchi

August 9th, 2023

UNIVERSIDAD DE COSTA RICA

The LHCb Online System



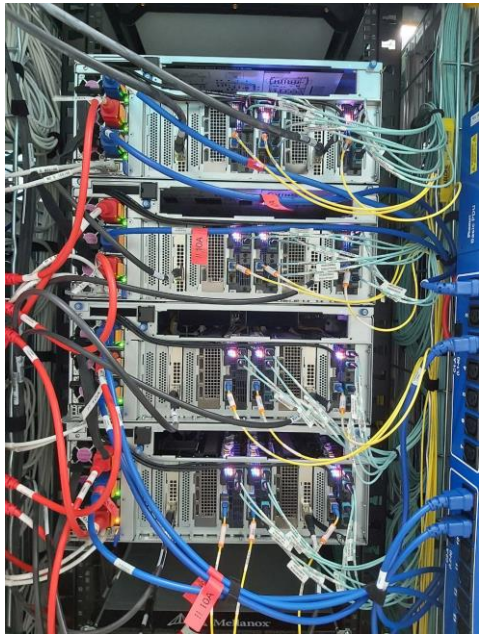
Trigger-less readout system

- Readout board (PCIe40)
- Event Builder
- HLT1 → storage buffer
- HLT2 → storage buffer
- Offline storage

Motivation

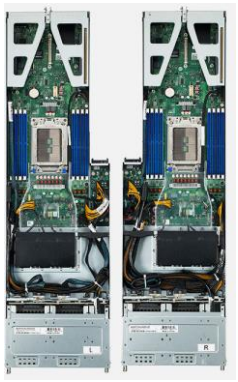
- **Future upgrades to LHCb datacenters**
  - ✓ Need to optimize energy efficiency
  - ✓ New constraints on operating conditions
- **Servers currently running on maximum performance mode**
  - ✓ Fans take up a large percentage of system power
  - ✓ More efficient configurations could highly impact power consumption
- **Evaluation of alternate cooling methods**
- **New servers for HLT2**
  - ✓ Establish a way to determine efficiency of systems along with performance metrics

Datacenters: ~ 2 MW



Event builder nodes

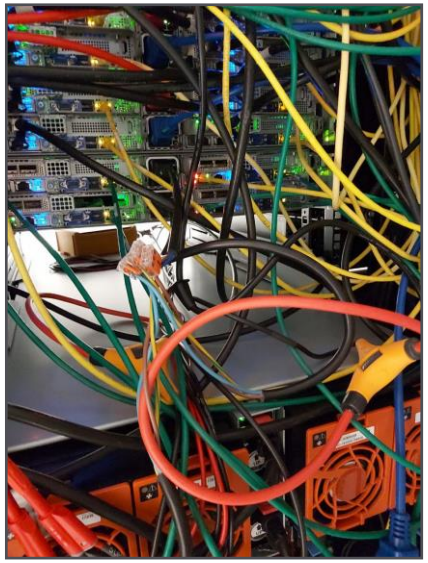
Supermicro Intel Xeon Gold Server



TYAN AMD EPYC Server

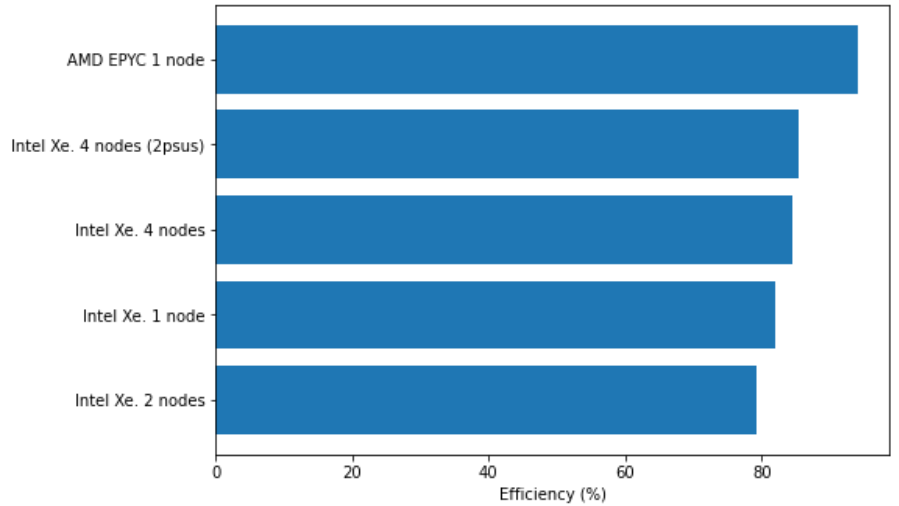
Efficiency and performance of servers

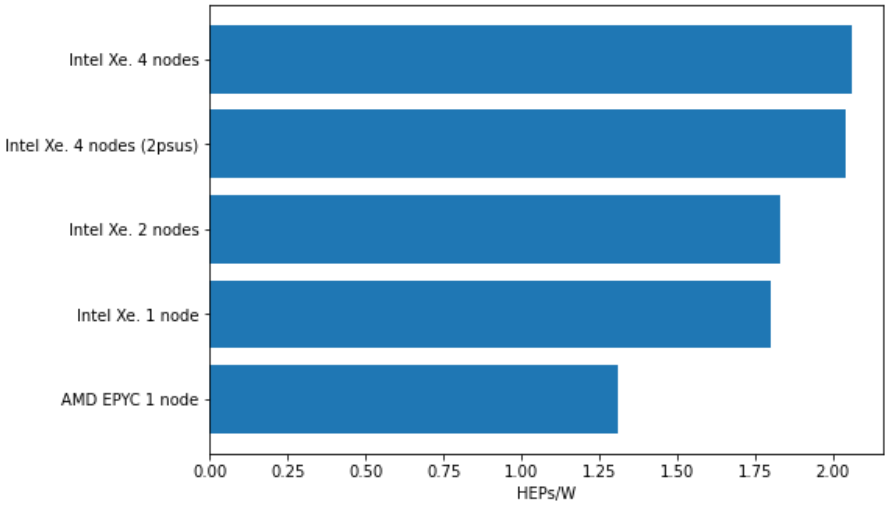
- **Power efficiency of Intel Xeon and AMD EPYC servers while running high CPU workload**
  - ✓ HEPscore23Beta benchmark
    - ✓ HEP workloads representative of computing usage of LHC experiments
- **Established metric to measure and compare performance: HEPscore/ W**
  - ✓ General purpose tool to select best platform CPU vs price vs power efficiency
- **Tested different generations of technology**
  - ✓ Difference in power consumption per memory module for DDR4 and DDR5



Efficiency and performance of servers

- **Power efficiency of Intel Xeon and AMD EPYC servers while running high CPU workload**
  - ✓ HEPscore23Beta benchmark
    - ✓ HEP workloads representative of computing usage of LHC experiments
- **Established metric to measure and compare performance: HEPscore/ W**
  - ✓ General purpose tool to select best platform CPU vs price vs power efficiency
- **Tested different generations of technology**
  - ✓ Difference in power consumption per memory module for DDR4 and DDR5



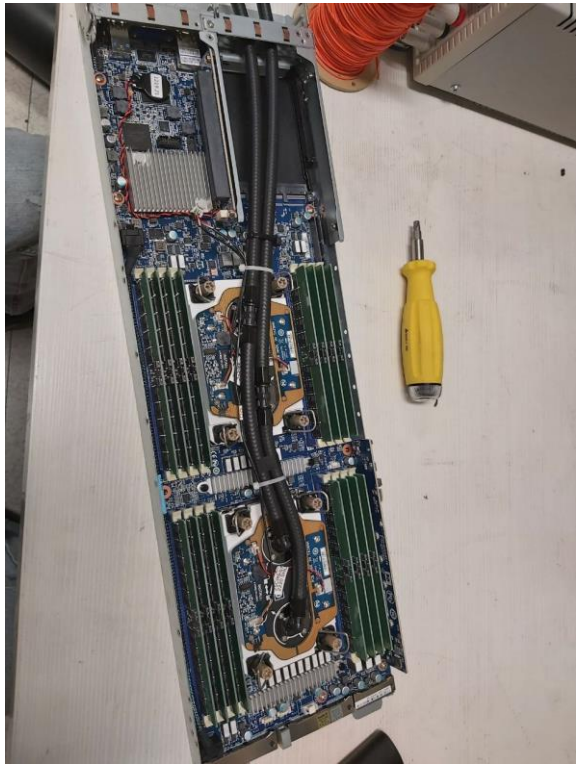


$$\frac{\Delta P_{\%RAMcapacity}}{\Delta\ modules} = 16.25\text{W/DDR4 module}$$

$$\frac{\Delta P_{\%RAMcapacity}}{\Delta\ modules} = 6.88\text{W/DDR5 module}$$

** Server with DDR5 has cold-plate cooling.*

Cold-plate Cooling Energy Efficiency



Results for server with cold-plate cooling:

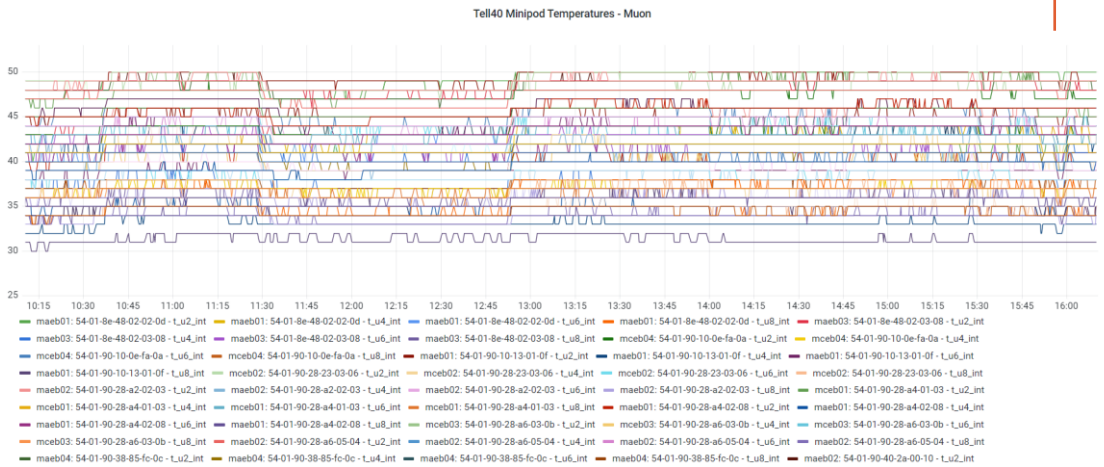
- **93.84% system efficiency**
- **2.97 HEPscore/W**
- Low fan activity
- **~ 79 W** for whole cooling system, which can be used for up to 40 kW of computing.

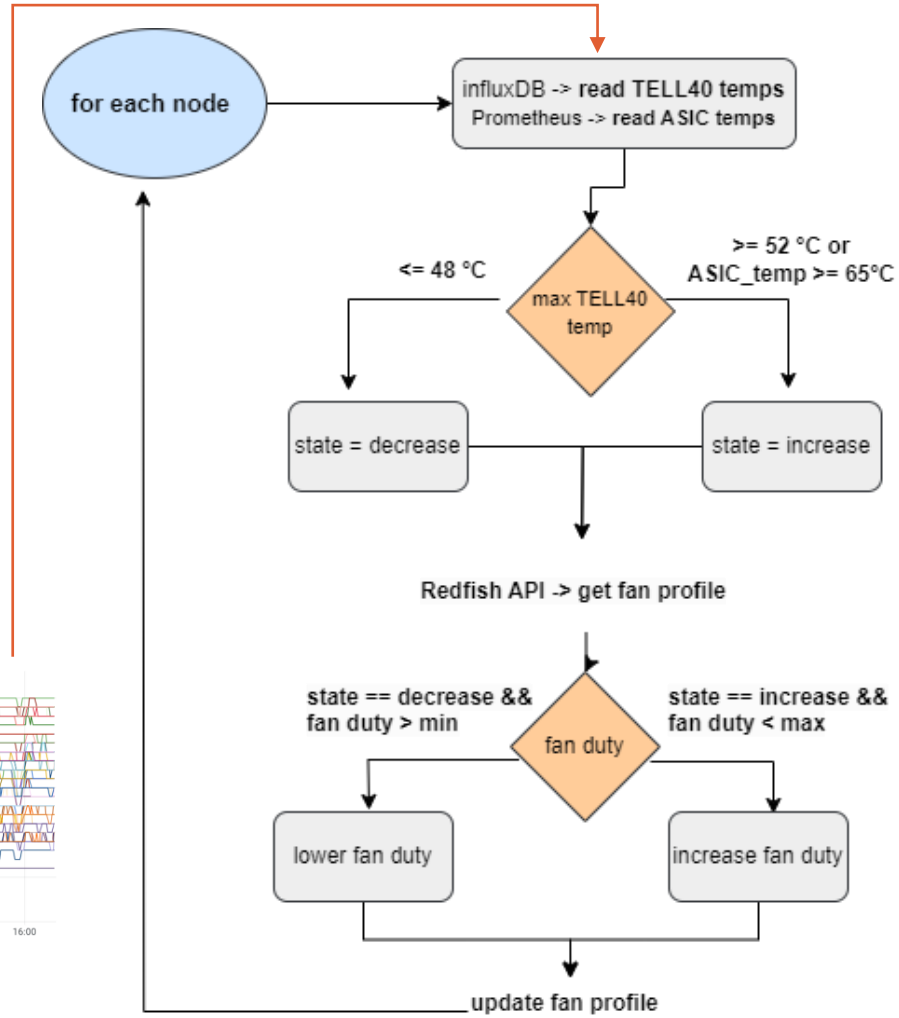
Supermicro Intel Quad server:

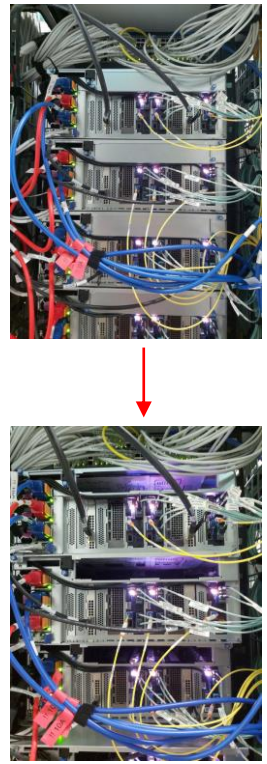
- 84.47% system efficiency
  - ✓ In contrast to rated power supply efficiency at the same load (96.22%): 11 % less, by not accounting fans
  - ✓ Translates to ~50 W per node due to fans (low estimate)
- For 40 kW of computing → **~ 4.4 kW** due to fan activity

Event builder fan control

- **PCIe40 card:** can work up to 70°C, for longevity it's best to keep it below 55°C. Workaround → **internal fan speed always at 100 %**
  - At idle, fan activity represented around 40% of system power.
- **Solution:** To control fan duty according to TELL40 transceiver temperature.
  - 2nd optimization: Remove metal covers to improve air flow.



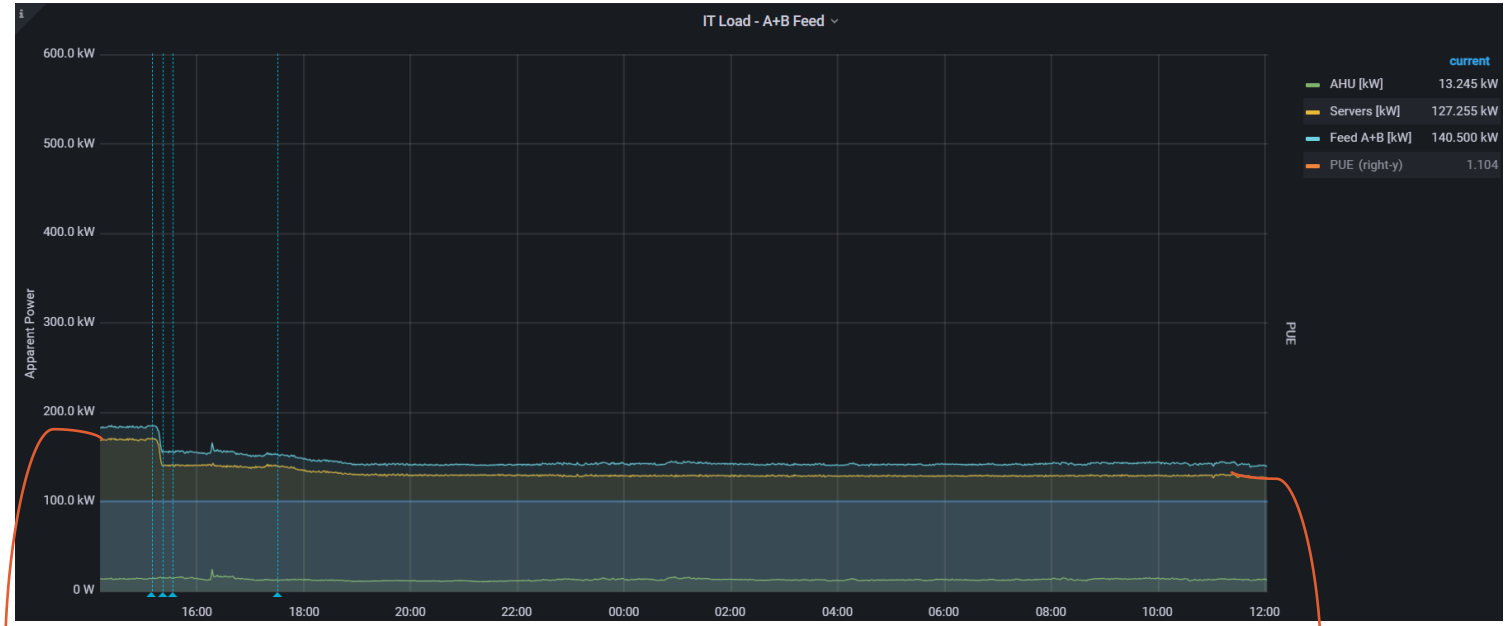




Event builder fan control

With the created fan duty control according to TELL40 transceiver temperature:

- **40 kW difference on IT3 container compared to previous fan configuration.**
  - ✓ 80 kW difference on two containers.
  - ✓ On one year → **700.8 MWh of energy saved.**



Previous configuration:
Fan duty at 100%
Servers: ~170 kW

New configuration: Fan duty based on TELL40 temp and removal of metal covers
Servers: ~129 kW

Conclusions

- Successfully measured power efficiency and quantified performance of different server configurations.
- Alternate cooling solutions, such as direct liquid cold-plate cooling, drastically reduce power consumption.
- Created a more efficient energy configuration by changing behavior of fans, which proved to save up significant power.

Further work

- Fine-tune fan control script with newly added ASIC temperature readings for optimized decision making.
- Analyze new fan configuration's impact on performance with reference workloads.
- Study impact and operation of new server designs based on acquired information.

Thank you

home.cern