# A new data-centre for the LHCb experiment

Loïc Brarda, Beat Jost, Daniel Lacarrère, Rolf Lindner, Niko Neufeld, Laurent Roy, Eric Thomas

Physics Department CERN CH-1211 Geneva 23, Switzerland

Computing in High Energy and Nuclear Physics, 2012

# Overview

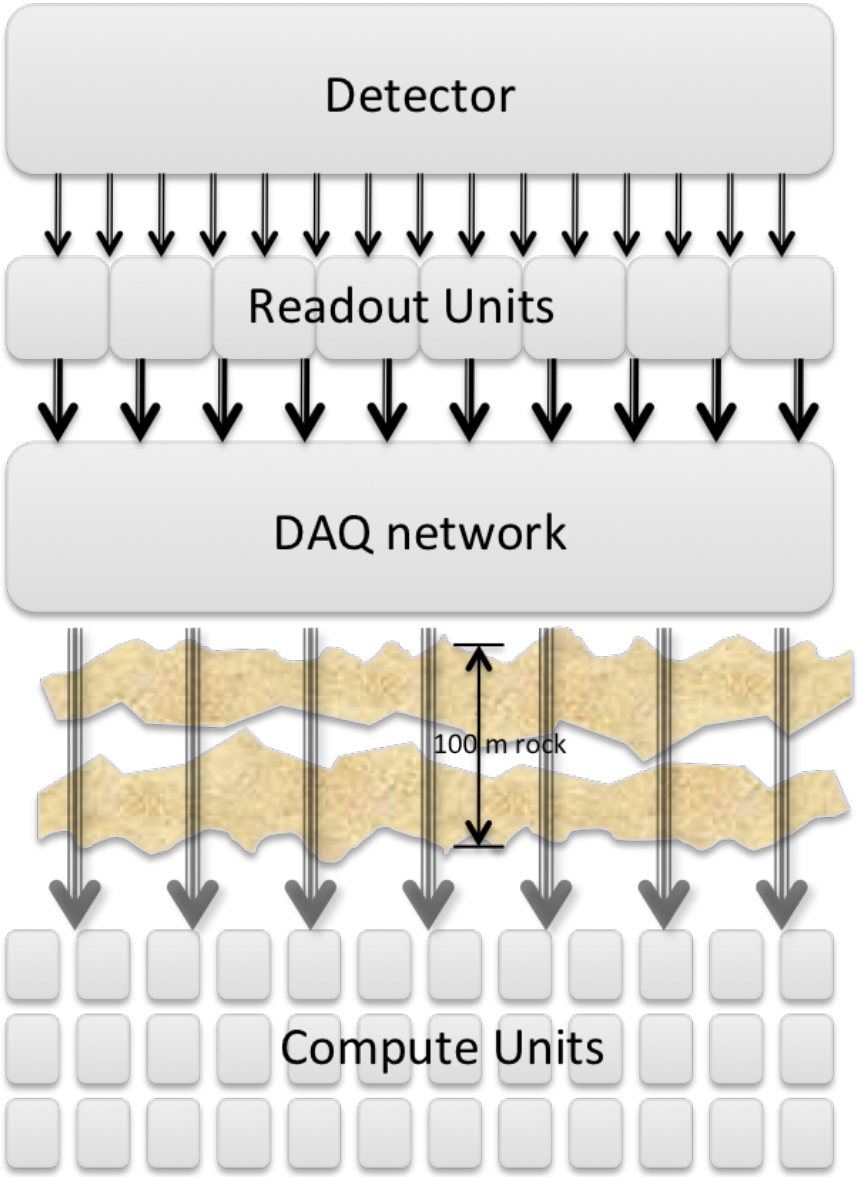
1. The LHCb upgrade, its DAQ and the requirements
   - DAQ
   - Farm and data-centre
2. Implementation Options
   - Brick and mortar
   - Remote hosting
   - Modular / Containerized data-centre

# The LHCb upgrade

- Major upgrade of the entire experiment during LS2 (2017 to 2018)
- Freely adjustable low-level trigger allowing readout speeds between 40 MHz and 1 MHz
- Allows to readout the entire detector at bunch-crossing rate (40 MHz)
- Sustain a luminosity up to $2 \times 10^{33}$

# Outline

# A 40 MHz DAQ




- GBT: custom radiation- hard link over MMF, 3.2 Gbit/s (about 10000)
- Input into DAQ network (10/40 Gigabit Ethernet or FDR IB) (1000 to 4000)
- Output from DAQ network into compute unit clusters (100 Gbit Ethernet / EDR IB) (200 to 400 links)

# Data Acquisition Requirements

| | |
|---|---|
| # of input links | 10000 |
| DAQ bandwidth per input link | 3.2Gbit/s |
| average total event-size | 100 kB |
| total bandwidth for the DAQ | 32 Tbit/s |
| output bandwidth | 2 Gigabyte/s |

- The data produced by a bunch-crossing in the collsion need to be "zero-suppressed" directly on the detector to reduce the number of input links from the detector.

# Outline

# Event Filter Farm requirements

- Capable of processing 30 Million events per second (out of 40 Million bunch-crossings)
- Scalable architecture (staged deployment)
- Houses up to 5000 dual-socket industry-standard servers
- Servers may be more general “Compute Units” (combination of servers and co-processor cards)

A “server” is understood to be a dual-socket machine, with at least 1 GB of RAM per hardware-thread

# Data centre requirements

| | |
|---|---|
| max # of servers | 5000 |
| max # of useful Us for servers | 2500 |
| number of Us for switches | 2 Us for 36 servers |
| Us for patch-panels per rack | 3 |
| depth of the racks | min 1000 mm |
| number of power feeds per rack | 2 |
| total usable IT power | 2 MW |
| minimum power per server | 350 W |

- low PUE
- long-term usefullness
- low cost (overall budget for the upgrade $\sim$ 50 MCHF)

# Non-requirements

- Safe power
  - Run servers from NFS root
  - Short processing time $\rightarrow$ neglible losses
  - Detector (data-source) not on safe power either
- Storage racks in the data-centre are not required
  - A small amount of onsite storage can be kept in an existing server-room
- Power-efficient CPUs $\rightarrow$ optimize CPU per unit of money only
  - CERN is a very large power-consumer so it gets good prices for electricity
  - We can go for very high power-density in a rack

# The current farm

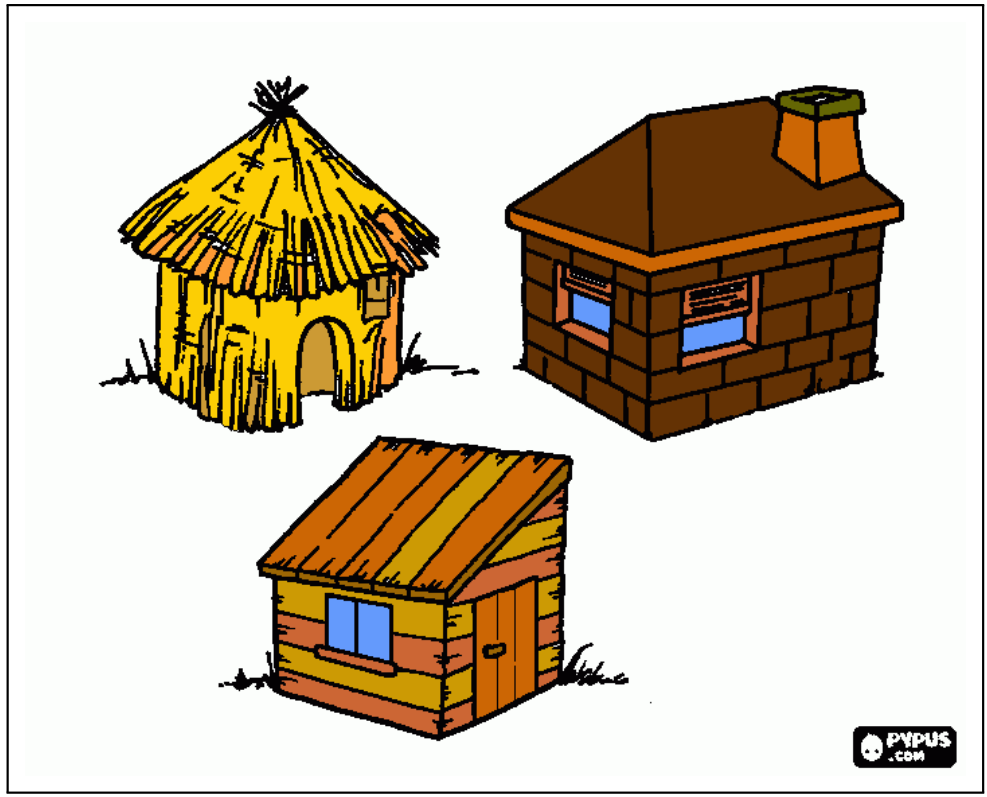
- Installed 100 m underground (UX85A)
- Accessible only by authorized personel under radiological supervision
- Reuses old detector-electronics racks (non standard posts, depth)
- Power and cooling limited to $\sim$ 500 kW

# The current farm 2

# Outline

# A brick-and-mortar data-centre

- Is said to cost 10 M$ per MW of IT capacity [1]
- This figure assumes a *commercial grade* facility: full battery backup, fly-wheels, redundancy, . . .



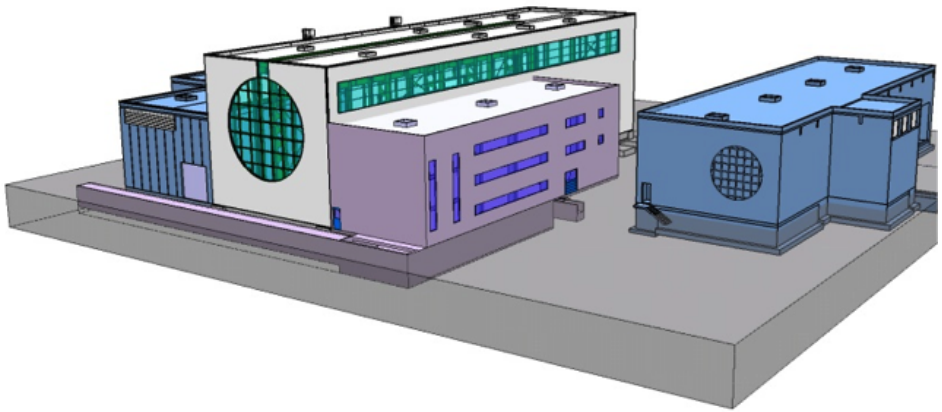
# (Cost-)factors in our favour

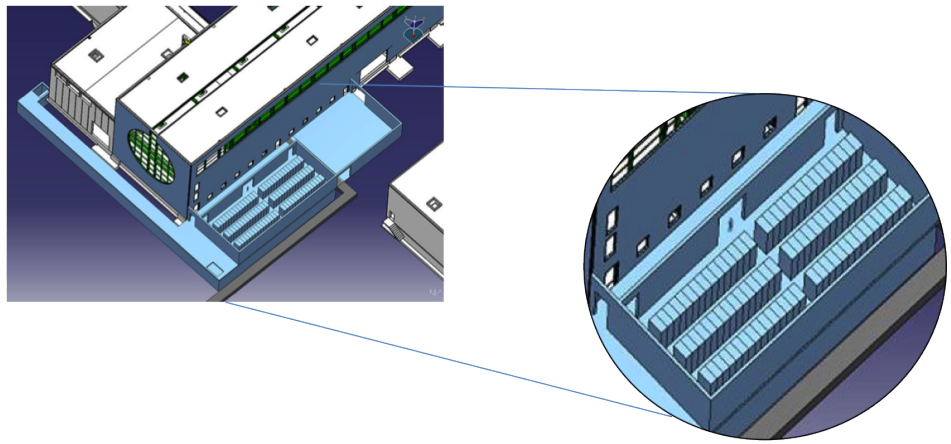
- Site available with high-power feeds and cooling plant
- Industrial site - only standard safety and environmental constraints $\rightarrow$ can be ugly and noisy as long as it is cheap
- Single-purpose ("customer") facility - no security sensitive data $\rightarrow$ only theft protection required

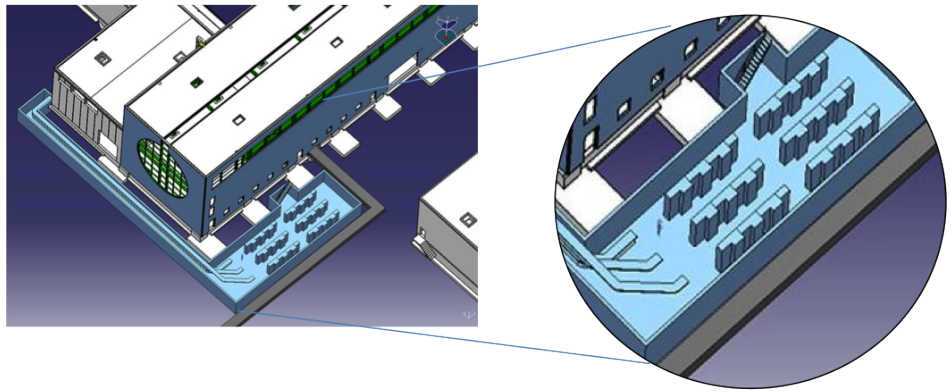
# The Point-8 site

# The first idea

- Collocated building with new control room and offices
- Utility feed from the basement
- Water-cooled racks (heat-exchanger doors)







# Outline

# The first suggestion from management: remote hosting

- Main argument: limited life-time of LHCb experiment (about 10 years after 2017). Long-term interest of CERN
- Minor argument: done by CERN/IT, use the "Cloud"
- Actually 3 options
  1. Remote hosting off the CERN site
  2. Remote hosting on the CERN site in an existing centre
  3. Remote hosting on the CERN site in a new centre $\rightarrow$ much more expensive than on-site - won't discuss further

# A technical aside: data-transport on a small number of fibres

- In both remote hosting options it is necessary to transport 32 Tbit/s over a very small number of fibres
- This is obvious for off-site hosting (cost of fibre / bandwidth rental)

Considerations are based on 10 Gbit/s and 100 Gbit/s technologies

- Minimal number of fibre-pairs using e.g. LR4 (25 Gbit/s signalling) 1280 if only a single wave-length is used
- 400 lambdas and more can be multiplexed on a single fibre

# Case-study on data-transport

A detailed study with one major vendor has been done, assuming an active solution (where the individual 10 Gigabit lanes are fully monitored and managed, as opposed to passive multiplexing)

- Assumed about 10 km distance, direct connection
- No cost for fibre laying or rental
- Cost per 10 Gigabit is found to be 7000 USD for 2016 (including an aggressive price-compression per year)
- This includes redundancy and a failver within a few minutes

Alternatives are using coloured interfaces at the output of the DAQ network (expensive) and passive multiplexing (DWDM) (difficult to manage)

- In any case privately such an infrastructure requires trained personel and has high maintenance costs $\rightarrow$ unrealistic in our environment

# Remote hosting off the CERN site

## Costs

- Renting the data-centre
- Power
- Renting of the data-path
- Data-transport

Last point alone is significantly above budget

# Remote hosting on the CERN site

Assume that a large part (about $\frac{2}{3}$) of the capacity of the computer centre 513 could be used

- Should fit easily within 10 km
- Can use multiplexing (expensive, see before) or install 10 Gigabit fibres (also expensive $\sim$ 12 MCHF according to a CERN internal estimate)

$\rightarrow$Seems expensive in comparison to the next option

# Outline

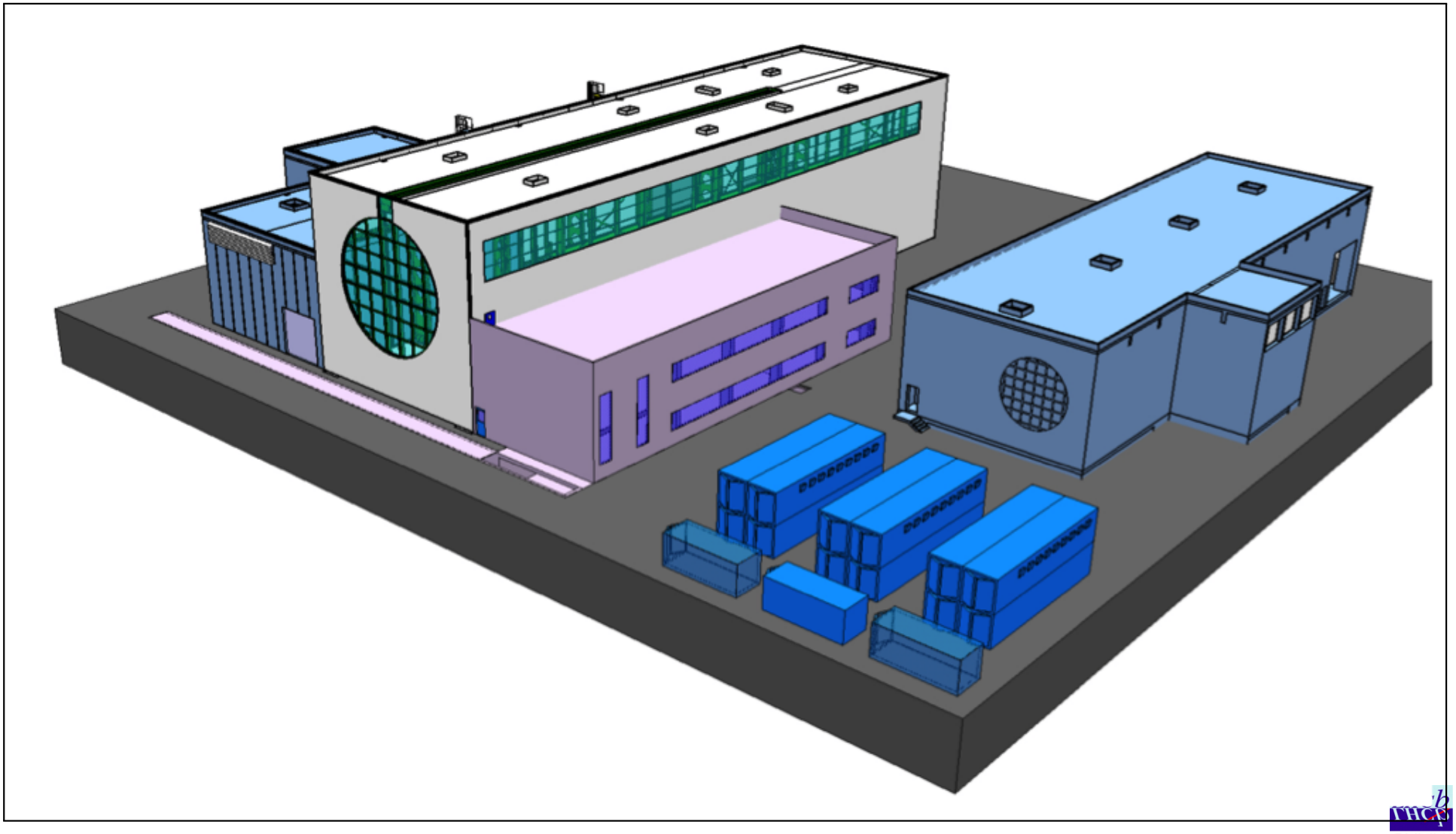
# The solution: modular data-centre

- Studying in detail various cooling solutions showed that to achieve a very low PUE $< 1.2$ the PUE must be the primary design criterion
- Constraints on building are conflict with co-use as office-building and control room hosting

# Advantages of containerized solution

- Lowest PUE (when using free cooling) - optimised design with integrated control
- Ease of re-use at the CERN site
- Capital investment costs lower or at least not higher than brick and mortar
- Could be re-sold if no more need
- Separation of IT infrastructure and office-usage
- Ready to use - requires only power (and a foundation)

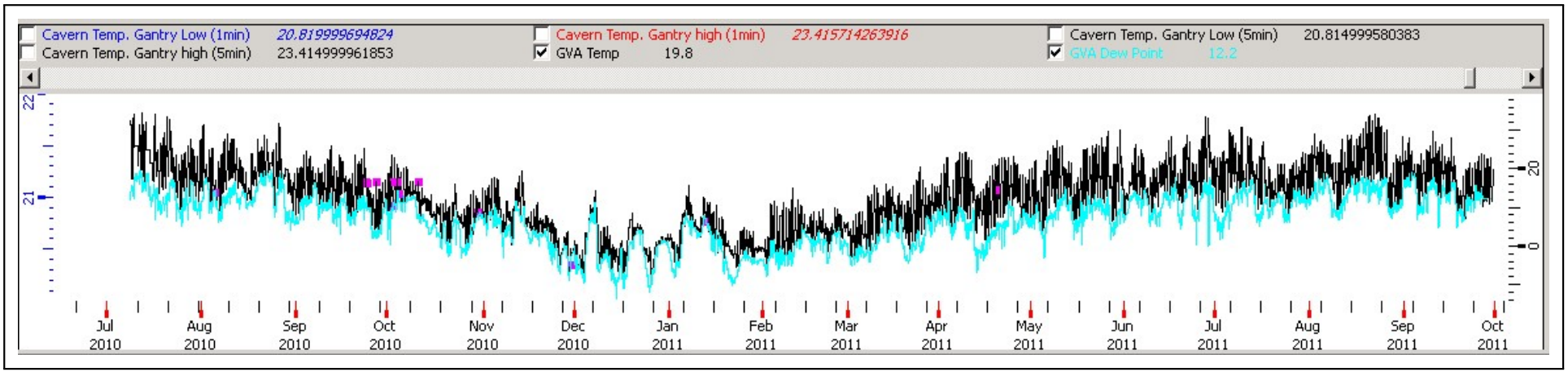
# How it could look like



# Results from prototype studies

From discussions with two major providers we established:

- the feasibility (budget)
- vendor neutrality (for the servers)
- most likely no need for additional cooling according to latest ASHRAE recommendations [2]



# Summary

- Taking advantage of the special requirements for the LHCb event-filter farm a cost-effective hosting in on-site containers has been identified
- Other options are more expensive and often suffer from the need of a long-distance transport of a large amount of data
- Such a solution will offer the best PUE and can most likely operate without any additional cooling

# What's next?

- Start construction of new control-room and office-building next year
- Transfer LHCb CR at the end of LS1
- Define final requirements for modular data-centre
- Competitive call for tender will be done in 2017

# For Further Reading I


M. Bramfitt and H. Coles.
Modular/Container Data Centers Procurement Guide: Optimizing for Energy Efficiency and Quick Deployment.
[Online]. Available: http://hightech.lbl.gov/documents/data_centers/modular-dc-procurement-guide.pdf, 2
2011.


the green grid.
Updated Air-Side Free Cooling Maps: The Impact of ASHRAE 2011 Allowable Ranges, 2012.
[Online]. http://www.thegreengrid.org/ /media/WhitePapers/WP46UpdatedAirsideFreeCoolingMapsTheImpactof