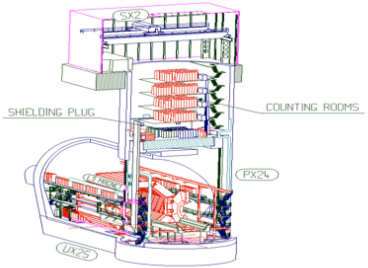
Run 3 data centre infrastructure for ALICE and LHCb

Joint workshop on DAQ, 13 April 2016

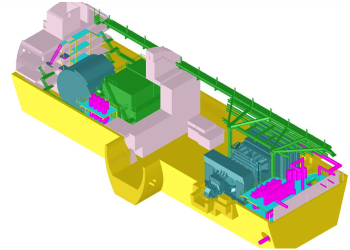
Existing installation ALICE

- DAQ and HLT are located in counting rooms on top of the PX24 shaft.



Existing installation LHCb

- HLT farm and DAQ boards are located underground (barracks on protected side of the LHCb cavern).
- $\sim$ 1700 HLT nodes, $<$ 11 kW per rack.



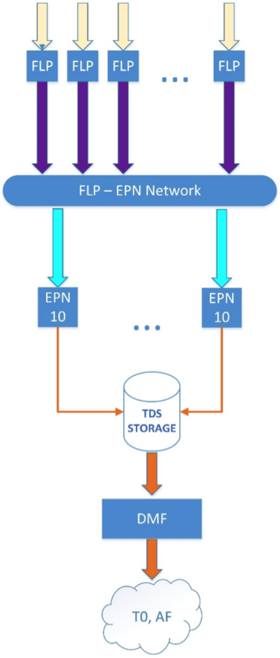
Limitations

- Legacy environment from LEP times.
- Limited room for growth.
  - Cannot fit more than 40 – 50 racks in existing barracks/counting rooms.
  - In case of ALICE there is also a weight issue (suspended counting rooms).
- Cooling based on active rear-door heat exchangers running with mixed water ($\sim 14^\circ$ C).
  - Not a "green" technology (chiller). Not easily scalable to power (density) needed for upgrade.

ALICE O^2

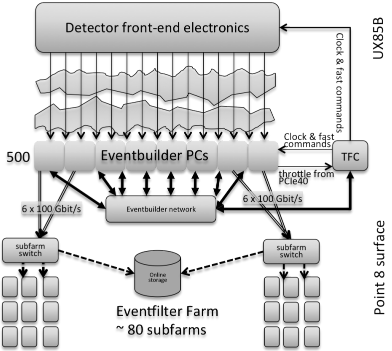
- First-level processors host readout cards and perform local data volume reduction.
  - 8000 optical links from detector front-ends.
  - Total bandwidth $\sim$ 10 Tbit / s (input).
  - Foreseen to stay in CR1 and use existing racks.
- Event-processing nodes reduce data volume by global reconstruction. All events go to storage.
  - Large installation. Requires a new room.
- Full system to be ready for start of Run 3.
- Deployment of prototype in 2017.

Component	IT Power [kVA]	Racks
FLP	213	17
EPN, storage	1937	45



LHCb online upgrade

- Event-builder servers house PCIe40 readout cards.
  - $\sim$ 15000 optical fibres from detector front-ends. Total bandwidth $\sim$ 40 Tbit / s.
  - Full system deployed from day 1 (commissioning during LS2).
- Event-filter CPU farm reduces data rate to storage to $\sim$ 100 kHz.
  - Foresee regular upgrades and extensions (don't need full capacity from day 1).



Component	Rack space [U]	IT Power [kVA]
Event-builder, ECS	1120	308
Event-filter (HLT)	2000	1400
Switches	544	109

Space and power requirements

- ALICE O^2 farm (without FLPs): about 2300 rack units and 2.1 MW power.
  - High power density due to heavy use of accelerators.
- LHCb upgrade farm (DAQ and HLT): about 3700 rack units and 2 MW power.
- No downtime during and high availability outside datataking.
  - $2N$ redundancy, dual feeds. Powered by same network as LHC machine.
  - Safe power (UPS, flywheel) would be nice to have but doesn't seem affordable for 2 MW.

Location, location, location

- Most economical solutions (both for ALICE and LHCb) are new data centres at Point 2 and Point 8 (surface).
  - Allows short distances for high-speed connections.
  - Use existing services.
- In theory, remote hosting would also be possible (except for event-builder).
  - But significant overhead due to infrastructure for long-distance data transport.
- Traditional brick-and-mortar implementation probably not a realistic option.

Energy efficiency

- In line with CERN's efforts to become more "green", aim at

$$\text{Power Usage Efficiency (PUE)} = \frac{\text{Total facility power}}{\text{IT power}} \sim 1.1 - 1.2.$$

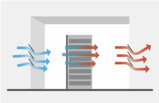
- Largely driven by cooling.

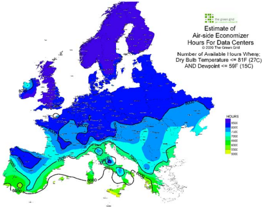
How to become "greener"?

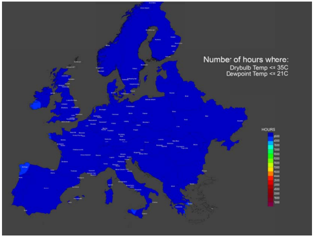
- Free cooling: use outside air (or water).
- Operating servers at elevated inlet temperature helps (in general).
    - ASHRAE 2011 allowable range (class A1 equipment): 15 to 32° C.
- Optimise air flow.
- Cool closer to the source of heat.
- Don't over-humidify/dehumidify.
- Use liquid cooling.
- Will discuss 2 1/2 cooling options in a bit more detail.
    - All compliant with target PUE.

Cooling solution 1: free (air) cooling

- Cooling without a compressor, using outdoor air (“opening the window”).
- Depending on environmental conditions, can achieve very low PUE.
  - In particular for direct free cooling (moving filtered outside air through the data centre).
  - Might even allow switching off server fans.
- During hot periods, use evaporative cooling or chiller.
- Geneva climate allows operation in free mode during most of the year.
- Well matched to containerised data centres.

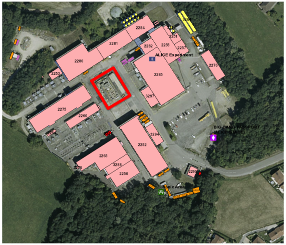






Modular (containerised) data centres

- Careful design and optimisation of air flow.
- Site preparations limited to providing foundation, electricity and water (if needed).
- Lead time about three months to half a year.
- Scalable, "buy as you grow". Could be reused later for a different project.
- Fire protection less critical than for buildings. Noise can be an issue though.
- Larger modules fit about 20 racks in a container. Typically high number of Us / m^2.



Point 2

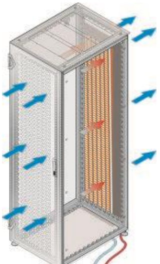
Point 8

Modular (containerised) data centres

- Ongoing market survey to identify manufacturers able to provide containerised data centres of the required scale based on either free air or water cooling.
- https://cds.cern.ch/record/2141082

Cooling solution 2: rear-door heat exchangers

- Water from cooling towers available “for free” at experimental sites.
  - Requires implementing a secondary circuit.
  - Temperature subject to seasonal variations. Expect max. temperature in secondary circuit (i. e. at rack inlet) $\sim 27^\circ$ C.
- Passive option: air is moved by server fans.
  - High reliability and low maintenance effort (no moving parts). No extra noise.
  - Equipment must provide good horizontal (front-to-back) air flow.
- Available cooling power depends on dimensions, flow rates, and temperatures.
  - For instance, with 800 mm $\times$ 47 U racks, water at 27° C ($\sim 3\,m^3\,/\,h$) could cool about 20 kW per rack while keeping room temperature at $< 35^\circ$ C
  - In theory, could go up to $\sim$ 40 kW (if using chilled water).
- Successfully implemented e. g. at GSI GreenCube and predecessors.
- Less demanding in terms of room layout. Compatible with container-based deployment and installation in a building.
  - For LHCb, installing the data centre in an existing assembly hall at Point 8 could be an option.



Cooling solution 2 1/2: liquid cooling

- Direct Liquid Cooling could be an interesting option.
  - Good match for warm (cooling-tower) water.
  - Needs complementary cooling to remove residual heat.
- Immersion cooling could be an option for HPC-like part of the farms.
  - Allows high power density and very low PUE. No noise.
  - Requires modified servers. Implications on connectivity and operation.