ATLAS DAQ/HLT Infrastructure

H.P. Beck[1], M. Dobson[3], Y. Ermoline[4], D. Francis[3], M. Joos[3], B. Martin[3], G. Unel[2], F. Wickens[5]

[1] Universität Bern, Sidlerstrasse 5, CH-3012 Bern Switzerland
[2] University of California, Irvine, Ca 92717-4575 USA
[3] CERN, CH-1211 Geneva 23 Switzerland
[4] Michigan State University, East Lansing, MI 48824-1321 USA
[5] Rutherford Appleton Laboratory, Didcot, Oxon OX11 0QX UK

Abstract

The ATLAS DAQ/HLT equipment is located in the underground counting room and in the surface building. The main active components are rack-mounted PC's and switches. The issues being resolved during the engineering design are powering and cooling of the DAQ/HLT equipment, monitoring of the environmental parameters, installation and maintenance procedures. This paper describes the ongoing activities and presents the proposed solutions.

I. DAQ/HLT OPERATIONAL INFRASTRUCTURE

The ATLAS DAQ/HLT system handles data coming in parallel from the detector. The readout sub-systems (ROSes), located in the underground counting room, and the computing farm (High Level Trigger and Event Builder), located in the surface building, are both based on rack-mounted PC's and network switches. About 2500 different components are will be housed in ~120 racks.

It is now entering the installation stage and therefore needs an operational infrastructure:

- counting room in SDX1 for DAQ/HLT equipment;
- housing and operation of rack-mounted equipment,
- monitoring of the environmental parameters,
- installation, operation and maintenance procedures.

The engineering design of the infrastructure is a common activity of ATLAS DAQ/HLT and Technical Coordination together with CERN Technical Support division. The task of the DAQ/HLT is to define their specific requirements and to find common solutions within DAQ/HLT and, where possible, with other experiments and Information Technology division.

A. Counting room

The 2 floor counting room in SDX1 building for housing HLT/DAQ equipment have been designed and built by CERN Technical Support division. The size of barrack is constrained by the crane, the shaft to USA15 and existing walls of the SDX building.

It accommodates 100 racks of 600 mm x 1000 mm and up to 52U high (height from the floor to the ceiling is ~2600 mm). The metallic structure was initially designed for 500 kg/rack load, but later re-designed for 800 kg/rack (electronics is heavy nowadays).

Lighting, ventilation, mixed water piping and cable trays were installed prior to the /HLT equipment installation and the temporary is provided for the time being (the final power distribution will include the supply from network and UPS).

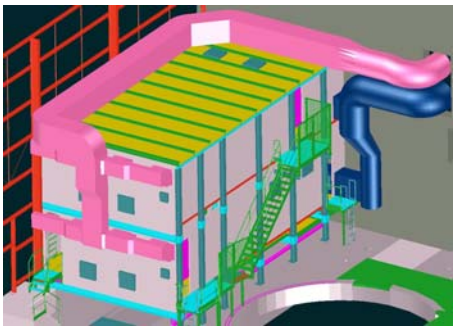

Figure 1: Counting room in SDX1

Air-conditioning, installed in the counting room, removes ~10% of the heat dissipated; other 90% are removed via water-cooling of the racks.

B. Housing of equipment

Two options were investigated during preliminary studies of the DAQ/HLT rack prototype [1]:

- modified “standard” 52U ATLAS rack for the ROSes in the underground counting room in USA15,
- industry standard Server Rack for SDX1 (e.g. RITTAL TS8).

For the racks in USA15 positions were already defined and their weight is not an issue. The difference form the “standard” 52U ATLAS rack is a horizontal air-flow cooling as the ROSes are based on the rack-mounted PC's.

Figure 2: DAQ/HLT racks in SDX1

For the racks in the SDX1 counting room there are height and weight limits which leaded to the lighter aluminium rack. The racks will be bayed in rows with partition panes for fire protection between racks. They will also exploit the horizontal air-flow cooling.

The PC's will be mounted from the front of the rack on supplied telescopic rails while switches will be mounted from the rear or front on support angles (if they are heavy). All cabling will be done at rear inside the rack.

C. Cooling of equipment

Common horizontal air-flow cooling solution has been adopted for both ATLAS "standard" and RITTAL TS8 racks. This is the outcome of joint "LHC PC Rack Cooling Project" project [2], which leads to a common water-cooling solution for the horizontal air flow inside the racks for.

The requirement was to evacuate ~10 kW of the power dissipated per rack by mean of the water-cooled heat exchanger fixed to the rear door of the rack.

The cooler from CIAT is mounted on the rear door (+150 mm to the depth of the rack). Its dimensions are 1800 x 300 x 150 mm^3, the cooling capacity is 9.5 kW and inlet water temperature is 15°C. It also has 3 fan rotation sensors for monitoring.

D. Powering of equipment

For the power distribution, the DAQ/HLT formulated their requirements, supplied the Technical Coordination with results of preliminary studies and working together towards the most cost effective final design.

The design and rating of the electrical distribution system for the rack mounted PC's requires a number of considerations such as inrush current, harmonic content of current and neutral loading:

- high inrush current may be handled by using D-curve breakers and sequential powering,
- harmonic content requires reinforced neutral conductor and breaker.

The solution, developed during prototype rack studies, was based on the assumption that 11 kVA of apparent power will be delivered per rack on 3 phases, 16A each. With typical PFC of 0.9 about 10 kW of real power will be available (and dissipated) per rack. In order to handle high inrush current, individually controllable breakers with D-curve will be used on each phase, which provides a first level of power sequencing. The 3 sequential power distribution units inside the rack (e.g. SEDAMU from PGEP) will provide a second level of power sequencing - 4 groups of 4-5 outlets with max 5A per group are powered-on with 200 ms separation in time.

There were also investigated possibilities for individual power control via IPMI or Ethernet controlled power units.

For the final implementation, the Technical Coordination proposed [3] to deliver two 3-phase cables to each rack and then provide a distribution to 6 single phase lines with 6-7 connectors on each line as shown in Figure 3.

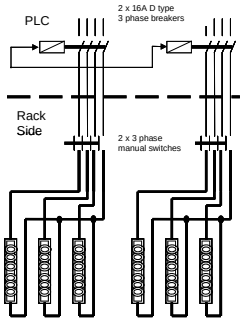

Figure 3: Power distribution inside the rack

There are 16A breakers on each individual phase, controlled from PLC. The installation is designed to sustain 35A of inrush current (peak) from 1 PC and for ~35-40 PCs per rack.

II. MONITORING OF OPERATIONAL INFRASTRUCTURE

The implementation of the monitoring of the environmental parameters is a full responsibility of the DAQ/HLT. We are aiming to have a single coherent management and monitoring tool based on IPMI, Linux tools and standard tools developed for the ATLAS Detector Control System and integrated into overall operation of the experiment.

The computing room environment is monitored by CERN infrastructure services (electricity, cooling, ventilation, safety) and by ATLAS central DCS (room temperature, etc.). The DAQ/HLT DCS will monitor rack parameters via two complementary paths - using available standard ATLAS DCS tools and using PC itself – lm_sensors and IPMI tools.

A. ATLAS DCS Tools

The following parameters will be monitored by the DAQ/HLT DCS inside the racks in USA15 and SDX1 computing rooms:

- air temperature will be monitored by 3 sensors (TF 25 NTC 10 kOhm from Quartz),
- inlet and outlet water temperature by sensor, located on the cooling water pipe,
- relative humidity – by HS2000V sensor from Precon,
- cooler's fan operation – by 3 sensors, provided by CIAT together with the cooler.

A status of the rear door (open/closed), water leak/condensation inside the cooler and smoke detection inside the rack will not be monitored by the DAQ/HLT DCS.

The "standard" ATLAS DCS toolkit is based on the ELMB - a general-purpose CANbus node for monitoring and control of sub-detector front-end equipment [4]. It is included as a component in a general CERN-wide control framework based on PVSSII SCADA system.

All sensors will be located on the rear door of the DAQ/HLT rack and sensor signals (temperature, rotation, humidity) and power lines will be routed to a connector on the read door to simplify assembly.

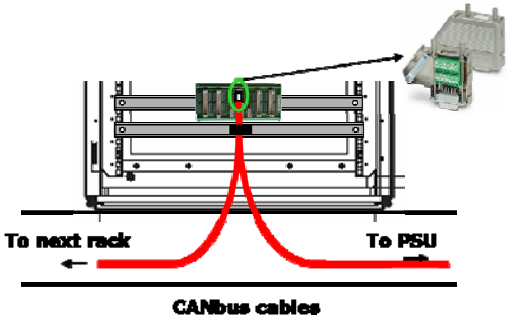

Figure 4: ELMB location inside the rack

Flat cable connects these signals to one of the ELMB motherboard connectors (3 connectors for 3 racks, 1 spare connector for upgrades). One ELMB is located at the rear of the rack (see Fugure 4) and may be used to monitor 3 racks.

B. *Use of internal PC's monitoring*

Most PC's nowadays come with a hardware monitoring chips (e.g. LM78) which proved measurements of onboard voltages, fan statuses, CPU/chassis temperature, etc. The Linux lm_sensors package includes a collection of modules for general hardware monitoring chips access and hardware monitoring. A small program, running on every PC, may use lm_sensors package in order to access CPUs temperature and other parameters and send this information. Interface to the control/monitoring PC running DCS application is achieved using DIM lightweight network communication layer which allows bi-directional communication.

An IPMI (Intelligent Platform Management Interface) specification is a platform management standard which intended to solve the problem of managing dissimilar hardware platforms. IPMI defines a standard methodology for accessing and controlling bare-metal hardware, even without software installed or running, effectively creating a standardized hardware management layer. Regardless of the nature of the underlying hardware, management software can use the one standard interface methodology to discover and communicate with a hardware platform, dynamically gather its relevant information and monitor its health and performance conditions.

The IPMI standard defines a hardware/software subsystem that performs key platform monitoring and management independently from the main processor. The specification is based on a specialized micro-controller, or Baseboard Management Controller (BMC), and defines a message based communication protocol for system management. The BMC operates on standby power and periodically polls system health variables such as temperature, fans, voltage, and power supplies and chassis intrusion. When discovering any anomaly, the BMC can log the event, send alerts via LAN and initiate recovery actions such as system reset.

In DAQ/HLT experience has been gained with IPMI v1.5. which allows access via LAN to reset, to turn PCs off & on, to login as from the console and to monitor all sensors (fan, temperature, voltages etc.). It can also log BIOS & hardware events.

III. INSTALLATION, OPERATION, MAITENANCE

Installation and maintenance procedure of the DAQ/HLT equipment is supported by the Rack Wizard (RW) - a graphical interface for electronics configuration and cabling databases. All DAQ/HLD equipment had been entered in the RW – rack positions and their content – PC's, switches, patch-panels, etc. After entering parameters, they have to be checked for consistency and maintained over the lifetime of the experiment.

A. *Pre-series installation*

The installation efforts now are concentrated on the physical installation of the pre-series components at CERN (~8% of the final size) - fully functional, small scale, version of the complete HLT/DAQ installed in 6 racks on the SDX1 lower level and USA15. This exercise shall highlight and solve procedural problems before we get involved in the much larger scale installation of the full implementation. The initial pre-series installation will grow in time – in 2006 there will be 4 more racks installed and in 2007 - 36 more racks. The full size of the system will be achieved in 2008.

The cabling of the DAQ/HLT equipment have beed given a special attention. All cables have been defined and labelled before and updated after installation.

Figure 5: Cabling inside the rack

Diring cable installation the attempt was made to minimize cabling between racks, to keep cabling tidy and to conform to minimum bend radii. Inside the rack it was decided not usi cable arms but uncable a unit before removal.

B. *Farm management*

System Management of HLT/DAQ computing farm has been considered by SysAdmin Task Force. Topics, addressed by the group, include users & authentication, networking in general, booting, OS, images, software and file systems, farm monitoring, how to switch on/off nodes, remote access & reset with IPMI.

The HLT/DAQ computing farm uses file servers/clients architecture, shown in Figure 6. It is connected to the CERN Public Network via gateway and firewall services. Tree servers/clients structure consists of a Central File Server, 6 Local File Servers and about 70 clients. All files come from a single (later a mirrored pair) Central File Server. All clients are net-booted and configured from Local File Servers (PC's & SBC's).

A maximum number of ~30 clients per boot-server allows scaling of up to 2500 nodes. Node configuration is a top-down and machine/detector/function specific.

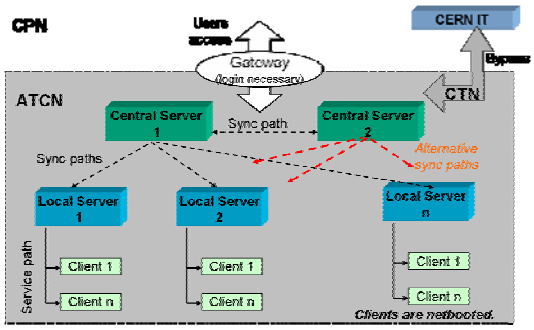

Figure 6: System architecture

Node logs are collected on the local servers for post-mortem analysis if needed.

All farm management related functions (lm_sensors, IPMI) are unified under a generic management tool – Nagios, which provides a unique tool to view the overall status, issue commands, etc. ATLAS DCS tools also will be integrated

IV. REFERENCES

[1] Y. Ermoline, B. Martin, F. Wickens, Prototype rack for computing farm, 30 September 2004.

[2] F. Wickens et al., The LHC PC Rack Project, 10th LECC Workshop, Boston, 13-17 September 2004.

[3] W. Iwanski, G. Blanchot, Power distribution in TDAQ racks, TDAQ coordination meeting, 10 May 2005.

[4] B. Hallgren et al., The Embedded Local Monitor Board (ELMB) in the LHC Front-end I/O Control System, 7th LECC Workshop, Stockholm, 10-14 September 2001.