7 Monitoring

7.1 Overview

Fast and efficient monitoring is essential during data-taking periods as well as during the commissioning of the detector. The quality of the data sent to permanent storage must be monitored continuously. Any malfunctioning part of the experiment must be identified as soon as possible. The types of monitoring information may be events, fragments of events, or any other kind of information (histograms, counters, status flags, etc.). These may come from the hardware, processors, or network elements, either directly or via the DCS. Some malfunctions can be detected by the observation of a single piece of information and could be performed at the source of the information. Infrastructure has to be provided to process the monitoring information and bring the result to the end user (normally the shift crew).

Collecting monitoring information can be done at different places in the DataFlow part of the TDAQ system: ROD crate, ROS, and SFI. Additional monitoring information can be provided by the LVL2 trigger and by the Event Filter since they will decode data, compute tracks and clusters. LVL2 and the EF will count relevant quantities for simple event statistics, for monitoring the functioning of the various trigger stages, and for tracking their selection power.

Monitoring information that is collected may be processed locally, e.g. in the ROD module, or in dedicated processes running in the Event Filter. Additional processing may be done by a dedicated Online Monitoring Farm possibly also used for calibration activities. Results will be sent to the shift crew in the control room.

It is clear that monitoring cannot be precisely defined at this stage of the development of the experiment and should therefore be kept as flexible as possible. The ideas presented in this section are the result of a discussion that has started recently at the level of the whole ATLAS community [7-1], [7-2]. The ideas are bound to evolve during the preparation of the experiment, as well as during its lifetime.

7.2 Monitoring sources

7.2.1 DAQ data flow monitoring

7.2.1.1 Front-end and ROD monitoring

Here, sub-detector front-end electronics and ROD module specific monitoring is addressed. It deals with:

- data integrity monitoring. Sampled event fragments must be collected at the level of the ROD and sent to dedicated monitoring processes to check the consistency of the data and the headers. Comparisons at the different levels of the transfers must be done. Use of the Data Collection network is preferred for this operation.

- operational monitoring. Throughput and similar information, scaler counters, histograms produced by the RODs must be transfered to the end user for further analysis via the control network.
- hardware monitoring, in cooperation with DCS. Additional monitoring of the hardware, not accessible to DCS (e.g. internal functioning of PCBs) may be provided at this level.

7.2.1.2 Data Collection monitoring

Here, it is DAQ specific monitoring which is addressed. The same points as the one listed in the previous sections will be addressed, at the level of the Data Collection (ROB, Event Builder, SFI, SFO):

- data integrity monitoring
- operational monitoring (throughput and similar, scaler histograms)
- hardware monitoring.

7.2.2 Trigger monitoring

7.2.2.1 Trigger decision

The general philosophy of the trigger decision monitoring is to re-evaluate the decision of each trigger level on samples of both accepted and rejected events in order to confirm the quality of the decision.

7.2.2.1.1 LVL1 decision

The LVL1 decision (for LVL1 accepts) is cross-checked when doing the LVL2 processing. In addition, a pre-scaled sample of minimum-bias LVL1 triggers will be passed to dedicated processing tasks (possibly in a dedicated partition of the Event Filter or in an Online Monitoring Farm).

7.2.2.1.2 LVL2 decision

The LVL2 decision (for LVL2 accepts) is cross-checked when doing the EF processing and pre-scaled samples of events rejected at LVL2 will be passed to the EF. Detailed information on the processing in LVL2 is appended to the event (via the pROS) for accepted and force-accepted events. This will be available at the EF for further analysis.

7.2.2.1.3 EF decision

Detailed information will be appended to the event, for a sub-set of accepted and rejected events for further offline analysis.

7.2.2.1.4 Classification monitoring

For monitoring, classification is a very important output of both LVL2 and EF processing. It consists of a 128-bit bitmap which records which signatures in the trigger menu were passed. Histograms will be filled locally on the processors where the selection is performed. With a maximum accept rate of 3.3 kHz for LVL2 and 200 Hz for EF, and assuming a sampling rate of 0.1 Hz, a 1 byte depth is sufficient for the histograms. For both LVL2 and EF farms, the bandwidth for the transfer of the histograms is therefore 3.5 kbyte/s.

7.2.2.1.5 Physics monitoring

In addition to classification monitoring, the simplest approach to monitoring the quality of the physics which is sent to permanent storage is to measure the rates for some physics channels. It can be performed easily in the EF. A part of the result of these monitoring operations will be appended to the event bytestream for offline analysis. Other data will be sent to the operator via the standard Online Software media for an online analysis and display.

An interesting approach to the physics monitoring could consist of a set of prescaled physics triggers with relaxed thresholds to monitor both the trigger decision and the effect of the trigger sharpness. Further studies will be performed to explore this approach.

Histograms of the rates for every item of the trigger menu as a function of time will be recorded, with the relevant variables with which they must be correlated (e.g. the instantaneous luminosity). Such histograms can very quickly give evidence of any malfunctions, although their interpretation may be quite tricky.

Well-known physics channels will be monitored so that one will permanently compare the observed rates with the expected ones. The list of such channels will be established in collaboration with the physics groups.

Information coming from the reconstruction algorithms executed for selection purposes in the EF may be of interest. One will monitor, for example, the number of tracks found in a given detector or the track quality at primary and secondary vertex on a per event basis. There again, a comparison with reference histograms may be of great help in detecting malfunctioning. Physics quantities will be monitored, e.g. mass-plots of W and Z, n-Jet rates, reconstructed vertex location, quality of muon-tracks, quality of calorimeter clusters, and quality of ID tracks. Input is required from offline reconstruction groups.

7.2.2.2 Operational monitoring

This section covers all aspects related to the ‘system’, e.g. the transportation of the events or event fragments, the usage of computing resources, etc.

7.2.2.2.1 LVL1 operational monitoring

The integrity and correct operation of the LVL1 trigger hardware will be monitored by processes running in trigger crate CPUs, and also by monitoring tasks in the Event Filter.

The LVL1 trigger is the only place where every bunch crossing is processed and where a crude picture of the beam conditions can be found. For example, the calorimeter trigger fills histograms in hardware of the trigger-tower hit rates and spectra. This is done for every trigger tower and can quickly identify hot channels which, if necessary, can be suppressed. Hardware monitoring is used to check the integrity of links between different parts of the LVL1 trigger, e.g. between the calorimeter trigger preprocessor and the cluster and jet/energy-sum processors. The Central Trigger Processor (CTP) will be monitored mainly at the ROD level, using internal scalers and histograms. It will include beam monitoring, including scaling of the trigger inputs on a bunch-to-bunch basis.

After the Event Builder, monitoring tasks, running in the Event filter or in a dedicated Monitoring Farm, will check for errors in the trigger processors on a sampling basis. This will be at a low rate, but will have high diagnostic power since details of the LVL1 caluclations can be checked. Event Filter tasks will also produce various histograms of trigger rates, their correlation, and history.

7.2.2.2.2 LVL2 operational monitoring

The LVL2 selection software runs within the framework provided by the Data Collection (DC) in the L2PU. It will therefore use the DC infrastructure and hence the monitoring tools foreseen for this system. The following relevant aspects of LVL2, will be monitored:

- trigger, data, and error rates
- CPU activity
- queue occupancies (load balancing).

Other valuable pieces of information for monitoring which could be performed by LVL2 processing units are:

- LVL2 selectivity per LVL1 trigger type
- number of RoIs per RoI type
- RoI occupancies per sub-detector
- RoI specific hit-maps per sub-detector.

In contrast to these data which do not require much CPU power, the monitoring of the quality of the data by LVL2 processors is not envisaged. Indeed, the available time budget is limited because of the necessity to release data from the ROB. Monitoring a fraction of the events in the L2PU is not desirable since this would introduce large variations in LVL2 latencies as well as possible points of weakness in the LVL2 system. The necessary monitoring of the LVL2 quality is therefore delegated to the downstream monitoring facilities, i.e. the EF (or online monitoring farm) and the offline analysis. One should however discuss very carefully the opportunity for the L2PU to fill some histograms, possibly read at the end of the run, so that high statistics information is given. This information can not reasonably be obtained by using events selected by forced accepts on a pre-sampled basis. The evaluation of the extra CPU load for such operations should be made.

7.2.2.2.3 EF operational monitoring

The monitoring of the data flow in the Event Filter will be done primarily at the level of the Event Filter data flow (EFD) process. Specific EFD tasks, part of the main data flow, will be in charge of producing relevant statistics in terms of throughput at the different levels of the data flow. They have no connection with other processes external to EFD. The detailed list of information of interest for the end user has not yet been finalized and will continue to evolve throughout the lifetime of the experiment.

Among the most obvious parameters which are going to be monitored, one might quote:

- the number of events entering the Farm
- the number of events entering each sub-farm
- the number of events entering each processing host
- the number of events entering each processing task
- the number of events selected by each processing task, as a function of the physics channels present in the trigger menu
- the same statistics as above at the level of the processing host, the sub-farm, and the Farm.

Other statistics may be of interest such as the size of the events, as a function of different parameters (the time, the luminosity of the beam, and the physics channel).

The results of the data flow monitoring will be sent to the operator via standard Online Software media (e.g. IS or Histogram Service in the present implementation).

7.2.2.2.4 PESA SW operational monitoring

A first list of parameters which could be monitored for debugging purposes and comparison with modelling results can be given:

- time spent in each algorithm
- frequency at which each algorithm is called
- number of steps in the step sequencer before rejection
- info and debug messages issued by the PESA SW
- number of active input/output trigger elements.

Some of these points could be monitored during normal data taking.

Profiling tools, such as NetLogger for coarse measurements and TAU, have already been studied in the context of LVL2, and their use on a larger scale will be considered by the PESA software team.

It is intended to make use of the ATHENA Histogramming service, which should therefore be interfaced to the EF infrastructure. Some control mechanisms should be provided to configure the various monitoring options and to operate on the histograms (e.g. to reset them after having been transferred).

7.2.3 Network monitoring

Monitoring the TDAQ network activity is essential during the data taking period. A very large number of switches (several hundreds) will be included in the network. Their management and monitoring is crucial. A first experience has been gained with the testbed used for network and data flow studies. Requirements are presented in [7-3].

Preliminary studies have been undertaken for tool evaluation. A first evaluation of the tool used by CERN-IT (*Spectrum*, from Aprisma) has been made. It has been found quite mature and it offers most of the expected functionality. Nice graphical interfaces are available giving an accurate and explicit view of the system (Figure 7-1). Extra work will be required to integrate the tool with the Online Software environment.

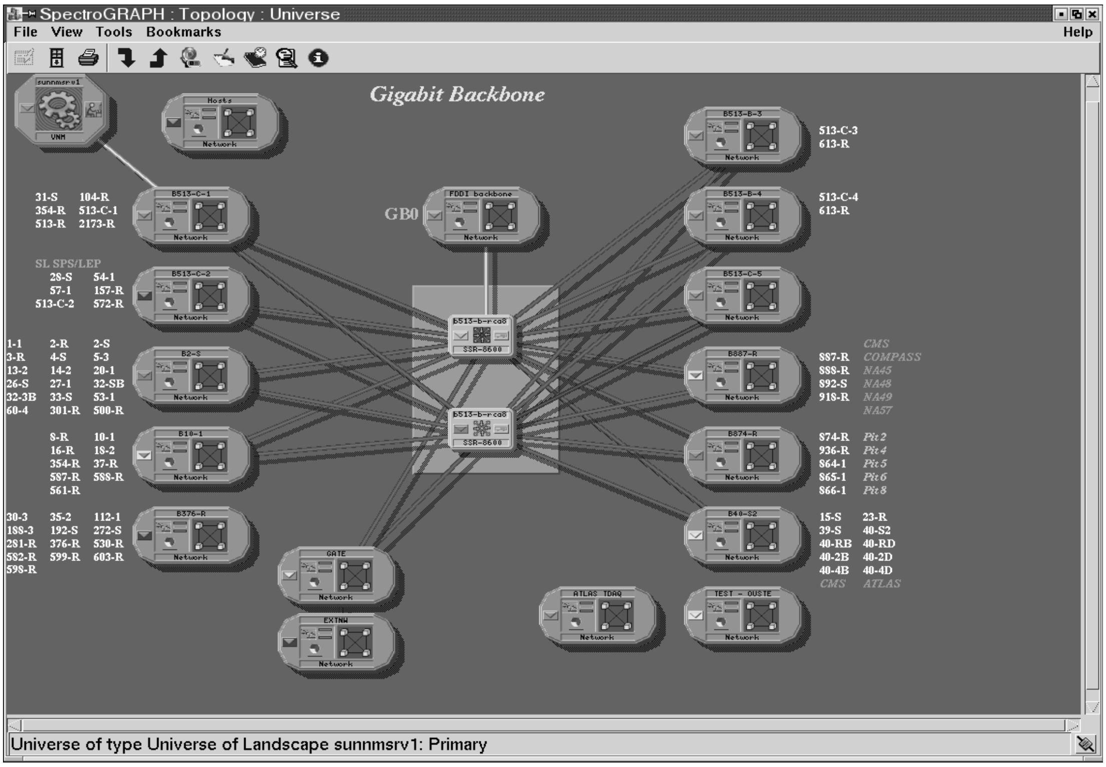

Figure 7-1 Example of a user interface of the Spectrum tool

7.2.4 Detector monitoring

The detector monitoring can be done at different places in the DataFlow part of the TDAQ system: ROD Crate, ROS, and SFI. Moreover, additional monitoring can be provided by the LVL2 trigger and by the Event Filter due to the fact that these programs will decode data, compute tracks and clusters, count relevant quantities for simple event statistics, and monitor the functioning of the various trigger levels and their selection power.

The ROD level is the first place where the monitoring of the data quality and integrity can be easily done. The computing power provided, for example, by the DSPs installed directly on the ROD board allows sophisticated calculations to be performed and histograms to be filled. Sending these histograms to analysis workstations will then be performed by the ROD crate CPU (using the Online Software tools running on this CPU).

Some detectors will need a systematic monitoring action at the beginning of the run to check the integrity of the system. This concept has already been introduced at the test beam by the Pixel sub-detector. At the beginning of the run and at the end there are special events with start and end-of-run statistics. The need for having this first check of the data at the ROD level is driven by the huge amount of information. If monitored later, the back tracking of possible problems would be complicated. The frequency of this activity, for normal operation, can be limited to the start and end of run.

Larger parts of the detector are available at the ROS level (compared to the ROD level), and monitoring at this level is therefore considered as a potentially interesting facility. A correlation between ROS crates is not seen as needed because such a correlation may be obtained at the SFI level. Event fragments sampled at the level of the ROS could be routed to dedicated workstations operated by the shift crew.

When information from several detectors is needed, the natural place to monitor it is after the Event Builder. The SFI is the first place were fully-assembled events are available. The monitoring at the level of the SFI is the first place in the data flow chain where the calorimeter, muon spectrometer, and LVL1 information is available. This can be used, for example, to cross-check the data for LVL1 calorimeter trigger towers against the more-precise, fine-granularity calorimeter readout. Moreover at the SFI level a first correlation among sub-detectors is possible and is seen as extremely useful.

When the monitoring requires some reconstruction operations, it seems natural to try to save computing resources by re-using the results obtained for selection purposes and therefore to execute this activity in the framework of the EF. In addition, some detectors plan to perform monitoring at the level of event decoding, i.e. in the bytestream conversion service, and to fill histograms during the reconstruction phase associated with the selection procedure in the EF. These histograms should be sent regularly to the shift operators and archived. More sophisticated monitoring operations might require longer execution times. However, since CPU power available in the EF should be kept in first priority for selection, it seems more efficient to have a dedicated monitoring farm running beside the EF.

Finally, some monitoring activities such as calibration and alignment checks may require events with a special topology selected at the level of the Event Filter. For instance, the Inner Detector group plans to perform the alignment of the tracking system online. This requires thousands of selected events, to be either stored on a local disk or fed directly to the processing task. Then CPU intensive calculations are required to invert matrices which may be as large as 30000×30000. With a cluster consisting of 16 PCs (as available in 2007, i.e. 8 GHz dual CPU PC with 1 Gbyte of fast memory and a 64-bit floating point arithmetic unit), this can be completed in less than one hour. A very efficient monitoring of the tracker alignment can therefore be performed. Similar requirements are made by the Muon Spectrometer for the purpose of monitoring and calibration operations.

7.3 Tools for monitoring

This section describes where and how (i.e. with which tools) monitoring operations will be performed.

7.3.1 Online Software services

The Online Software (see Chapter 10) provides a number of services which can be used as a monitoring mechanism which is independent of the main data flow stream. The main responsibility of these services is to transport the monitoring data requests from the monitoring destinations to the monitoring sources, and to transport the monitoring data back from the sources to the destinations.

There are four services provided for different types of monitoring information:

- **Event Monitoring Service** — responsible for the transportation of physics events or event fragments sampled from well-defined points in the data flow chain and delivered to analysis applications. These will examine and analyse the data in order to monitor the state of the data acquisition and the quality of the physics data in the experiment.
- **Information Service** — responsible for the exchange of user-defined information between TDAQ applications and is aimed at being used for the operational monitoring. It can be used to monitor the status and various statistics data of the TDAQ sub-systems and their hardware or software elements.
- **Histogramming Service** — a specialisation of the Information Service with the aim of transporting histograms. It accept several commonly used histogram formats (ROOT histograms for example) as the type of information which can be sent from the monitoring sources to the destinations.
- **Error Reporting Service** — provides transportation of the error messages from the software applications which detect these errors to the applications which are responsible for their monitoring and handling.

Each service offers the most appropriate and efficient functionality for a given monitoring data type and provides specific interfaces for both monitoring sources and destinations.

7.3.2 Computing resources for monitoring

7.3.2.1 Workstations in SCX1

It is foreseen to have several workstations in the the SCX1 Control Room near the experiment pit. These workstations will be owned and operated by the sub-detector crews who are on shift. They will receive, via the Ethernet network, the results coming from operations performed in ROD and ROS crates as well as event fragments. Whether the network is a dedicated one or the general-purpose one is still an open question. These workstations will perform subsequent treatment such as histogram merging or event display. They may possibly delegate processing to machines in remote sites if available local CPU power is not sufficient. Results of monitoring operations will be made available to the whole shift crew using the dedicated Online Software services.

7.3.2.2 Monitoring in the Event Filter

From the beginning of the design of the EF, it has been foreseen to perform some monitoring activities there, in addition to the ones related directly to the selection operation. EF is indeed the first place in the data-selection chain where the full information about the events is available.

Decisions from the previous levels of the trigger system can be checked for accepted events and for those that would normally have been rejected, but were retained on the basis of a forced accept. Information coming from the reconstruction phase, which generally requires a large amount of CPU power, can rather easily be re-used, leading to large savings in terms of computing resources.

Monitoring in the EF can be performed in different places:

- directly in the filtering tasks (which raises the problem of the robustness of the monitoring code),
- in dedicated monitoring tasks running in the context of the Event Filter (here one should think of passing the information gathered in the filtering task to take profit of the CPU power already used).

As already stated, the first priority of the EF must be the selection procedure which should not be jeopardized by introducing some CPU-intensive applications in the processing host. Monitoring in the EF should therefore be reserved to light-weight applications which would profit most from re-using pieces of information produced by EF Processing Tasks.

7.3.2.3 Monitoring after the Event Filter

In order to perform the CPU-intensive monitoring activities such as the ones described at the end of Section 7.2.4, a dedicated Online Farm should be provided. Such a farm is also necessary to perform the various calibration tasks which do not require the full offline framework. It would be fed by events specially selected in the EF, as well as directly by the general DataFlow through one or more dedicated SFIs (so that it may receive events selected at previous stages of the data acquisition chain). Such specially selected events may be events rejected at LVL1 or LVL2 (on a prescaled basis), or events tagged for calibration purposes (physical as well as non physical events, e.g. generated by a pulser or corresponding to empty bunches).

If such an Online Farm were to be used, one would require that the Data Flow Manager be able to route events towards specific SFIs according to a tag set at various levels of the data acquisition chain (front end, LVL1, or LVL2). The DataFlow framework developed for the Event Filter seems to be well suited for the distribution of the events to the different applications. Moreover, a uniform approach for the EF and the Online Farm would bring some flexibility for the global computing power usage, since intensive monitoring is likely to be more required during commissioning or debugging phases while physics quality is not the first priority, and conversely. The size of this Online Farm is still to be evaluated.

7.3.2.4 Network requirements

7.3.2.5 A monitoring matrix is presented in **Network requirements**

A monitoring matrix is presented in Chapter 2.

7.4 Archiving monitoring data

Data which are produced by monitoring activities should be archived by some bookkeeping service so that they can be cross-checked offline with more detailed analysis. One must also store events whose acceptance has been forced at any level of the selection chain. These events are necessary to evaluate precisely the acceptance of the trigger. Details of the archiving procedure will be established in cooperation with the Offline Computing group.

7.5 References

7-1 B. Di Girolamo et al., *Introduction to Monitoring in TDAQ*, ATL-D-OR-001, https://edms.cern.ch/document/382428/1

7-2 B. Di Girolamo et al., *ATLAS Monitoring Requirements*, in preparation

7-3 M. Zurek, *Network Management Requirements*, https://edms.cern.ch/document/391577/0.7