REAL-TIME DATA ARCHIVING FOR GTA *

R. A. Cole and W. H. Atkins
Los Alamos National Laboratory
P.O. Box 1663, Los Alamos NM 87545

Abstract

The architecture of the GTA control system, the nature of a typical GTA commissioning activity, and the varied interests of those analyzing the data make it challenging to develop a general-purpose scheme for archiving data and making the data available to those who will use it. Addressing the needs of those who develop and trouble-shoot hardware and software increases the challenge.

This paper describes the aspects of GTA that affect archiving operations and discusses how the features of the EPICS archiving module meet a variety of needs for storing and accessing data.

Introduction

At the Ground Test Accelerator (GTA), much of the data acquisition and control uses the Experimental Physics and Industrial Control System (EPICS) [1]. The actual data acquisition and control occur on a distributed network of Input/Output Controllers (IOCs), each running EPICS. Each IOC has its set of hardware input and output channels as well as a set of 'soft' channels. The number of channels in the various IOCs ranges from a few hundred to many hundreds. The sheer number of channels is one of the first challenges to archiving at GTA.

Only a small fraction of the channels in an IOC are useful for analyzing the physics of GTA. In some cases, there are several channels that appear to provide a needed piece of information. This usually happens when 'soft' channels are used to alter the raw data in some way—by smoothing, averaging, or applying an offset, for example. In these cases, a controls engineer is needed to specify which should be used; an incorrect choice by a casual user would invalidate the analysis of an experiment.

Data can be requested from EPICS in several ways. The two of most interest for archiving are 'on each system trigger' and 'whenever the value changes.' The first method is needed for channels associated with beam-related information, such as current or position. The second method is useful for other channels, such as those that measure temperature or pressure. With a beam repetition rate of 5 hz, archiving a large number of beam-synchronous channels for an extended period of time produces a very large volume of data.

The last item on the list of challenges to archiving is the isolated nature of the GTA network. The IOCs are on a classified network that is separate from the unclassified network accessible from most users' offices. Archived data resides on the classified network in files on Sun workstations. Most of those who analyze archived data work on Macintosh computers attached to the unclassified network. In order to let these people concentrate on physics issues, an effective, painless way is needed to transport the data between the two networks and transform the data into a form accessible to the Macintosh's.

A brute-force approach to archiving would just save everything that is available. For GTA, this would quickly overwhelm both the network and disk space, as well as being a burden to those who are trying to find data for a particular set of channels for a particular time period. In addition, this approach has the risk that a particular channel will be used for inappropriate purposes. And, finally, the mountain of data would bog down most mechanisms for moving data between the two networks.

The evolution of the EPICS archiving module, and the evolution of how it is applied to GTA's needs, have addressed many of the challenges presented above. The objectives have been to minimize the impact on network loading and disk space utilization and to minimize the need for 'guru' involvement in using the archiving software.

Commissioning

GTA commissioning activities are often focused on succinct questions, such as "How does RFQ position affect beam position?" Typically, one or several parameters will be changed and then a 'synchronous snapshot' will be initiated to archive a few dozen samples. The samples in a synchronous snapshot are correlated by time. The advantage in archiving such a snapshot is two-fold: occasional erratic data can be filtered out; and variations in data behavior can be measured and studied. The product of this sampling process is a file containing a set of synchronous snapshots. In most cases, analysis focuses on the mean and standard deviation for each snapshot; occasionally, a study requires the actual data.

Archiving real-time data

Prior to a commissioning activity, the commissioning team addresses several issues:

1) What channels should be archived? In many cases, a generic list of channels will be used. For special commissioning activities, a special list of channels may be developed. When a new module is being commissioned, the channel name for an item may have changed, which requires a change in the list.

2) How many samples should be in a snapshot? Often, a snapshot is considered to be a fixed number of samples.

* Work supported and funded by the US Department of Defense, Army Strategic Defense Command, under the auspices of the US Department of Energy.

Some other number is sometimes used. Occasionally a snapshot is based on time, such as "take however many good samples are obtained in 30 seconds."

3) What special restrictions should be applied to samples? Typically, a sample for a beam pulse is required to have a value for each channel; if one or more channels didn't report a value for the beam pulse, then the sample is discarded. Other conditions can also be applied. For example, if the beam current is below a threshold on a particular beam pulse, then the sample for that beam pulse can be discarded.

These criteria, and others, are stored in the file that guides the archiver. This control file serves as a permanent record of what channels were archived and what criteria were used in the archiving. It and the archive data file are kept together until they are written to tape and then deleted from the disk.

When the GTA operator initiates a snapshot with a mouse click, the archiver begins accumulating synchronous samples, subject to any restrictions specified in the control file. Since the IOCs are on a network, the arrival times of data at the operator workstation are unpredictable; the archiver copes with this condition to accomplish time correlation, based on the time-stamp attached to the data in the IOC [2]. If any temperatures, pressures, or other slowly varying parameters are being archived, the archiver incorporates these values in a special way into the synchronous samples. When the snapshot is complete, the operator is informed of the number of samples and the time range of the data.

Retrieving archived data

Following the commissioning activity, the control file and the data file are copied to tape and subsequently loaded onto disk on the unclassified network. The retriever, which extracts information from the data file, is most often used just to read the values for a small subset of channels and write the values to a text file that can be read by Microsoft Excel on Macintosh workstations.

This 'export' process is controlled by a script file that is part of the larger script that is used for loading the tape onto disk. Each physicist can have an individually tailored script that is automatically executed when the tape is loaded. This procedure makes the data easily available for analysis, without the analyst needing to take any special action.

A choice is available for exporting either the actual data or snapshot means and standard deviations. Using the latter produces smaller files, which are quicker to upload into the Macintosh and which don't use as much memory and other workstation resources. Some analysts choose other tools to do their analysis—these can also read the exported text files.

The retriever program, which runs on a Sun workstation, can also perform other functions. It can print data in textual form and plot it in a variety of ways. The printing and plotting can be for either raw data or for means and standard deviations. In some simple cases, all of the desired processing can be accomplished with the retriever.

As information is gained about the data, that information can be stored in the control file as a permanent record associated directly with the data. For example, analysis might show that an error has occurred in surveying the location of a beam position monitor. Comments can then be entered in the file indicating that an offset should be applied to beam position values. The retriever supports viewing, editing, and printing the control file, with the comments that have been added.

Development and Trouble-shooting

Both the archiver and the retriever are used in other contexts than commissioning activities. The software is useful for checkout during the development of both hardware subsystems and software programs. The archiving software has also been valuable for trouble-shooting activities.

For some activities, synchronous snapshots aren't appropriate for archiving. For example, data from some of the channels (such as temperatures and pressures) may be intermittent, rather than on every system trigger. For these cases, each channel is archived as an independent stream of values, saving each change in value. With this approach, the data can be scrutinized later to understand how the channels are actually behaving.

Intermittent problems can be diagnosed by having the archiver trigger the sampling process based on some data condition. The archiver holds incoming data in a 'pipeline', so that it's possible to store data both prior to and following the trigger condition.

For either approach, the retriever supports printing and plotting data, as well as exporting the data for use by other programs.

Utility logging

Some GTA hardware subsystems are monitored around the clock as a routine part of operations. The vacuum subsystem and cryogenic subsystem are two examples. For this purpose, the archiver periodically samples the channels of interest; ten minutes is a commonly used sampling interval.

Application programs

The design and implementation of the EPICS archiving software have placed a high priority on modularity and code reusability. Other programs can use selected parts of the archiving software without the need to 'reinvent the wheel.'

For example, the EPICS operator interface software uses parts of the archiving software to acquire synchronous data and plot it on the screen.

As another example, the GTA application program that automates the transmission scan of the RFQ uses the archiving software to save the transmission scan data on the disk in a form that can be accessed by the archive retriever. Not only does this exploit reuse of the archiving code, but it also allows unifying the procedures for handling experimental data—the same procedures used for archiver data also allow making data from applications programs available to experimenters who use Macintosh workstations.

Conclusions

The unparalleled state of the art capabilities of EPICS [1] provide a superb foundation for data archiving. These capabilities include EPICS' software bus (Channel Access), its distributed database architecture, and its modular architecture. These features directly support the implementation of programs, such as the EPICS archiver, which operate in the extremely powerful client/server mode.

The EPICS archiving module is the basis for a robust system for handling GTA data. Through the use of UNIX scripts, the needs of experimenters, developers, and operations personnel are served. The focus of the archiving software is on power, efficiency, and flexibility in saving real-time data to disk and in later retrieving data from disk. The rudimentary processing capabilities are adequate for many needs; where more involved processing is required, or where publication quality displays are needed, support is provided for exporting data to more sophisticated processing tools. With this approach, the investment in the archiving software is lessened, and data analysts are enabled to use the tools that are specially tailored to their analyses.

Appendix A

Archiver features

1) handles numeric, discrete, text, and waveform channels

2) takes samples as data streams, as synchronous snapshots, or as periodic 'grab samples'

3) sampling can be triggered by operator demand, on a data event, or periodically

4) for synchronous snapshots with sampling triggered by a data event, archiving of data can begin prior to the event

5) for synchronous snapshots, assembling time-correlated samples compensates for variations in arrival time; also compensates for slightly different notions of time-of-day in the IOCs involved; non-synchronous channels can be specified to receive special treatment to simulate synchronous channel behavior

6) for synchronous snapshots, samples can be included or excluded from synchronous snapshots depending on the values of channels in the sample

7) data can be stored in either a binary form that optimizes disk and network utilization or in a text form that provides very high immunity from network or computer crashes

8) supports checking the quality of synchronous snapshots, especially with regard to detecting channels that aren't beam synchronous

9) can be operated in interactive mode, script mode, or as a server that starts running when a workstation boots

10) in all modes of operation, setup for archiving is obtained from an archive control file; in interactive mode, directly supports editing and testing of the control file

Archive retriever features

1) retrieves data from any of the forms produced by the archiver; retrieval can be 'on the fly' to allow viewing recently archived data, without needing to stop the archiver

2) can also retrieve real-time data directly, without needing to go through an archiving step

3) allows specifying the channels to retrieve, the file position to start retrieving, and number of samples to retrieve

4) a retrieval always produces time-correlated samples, regardless of the source of the data; slight variations among IOCs' notion of time-of-day are accomodated; non-synchronous channels can be specified for special treatment to simulate synchronous channel behavior

5) printing, plotting, and exporting can use either the raw data or means and standard deviations for snapshots

6) prints data in textual form, with alarm status for each point; width of column, number of columns per line, and font size can be specified; if channel name exceeds column width, it is folded and printed in several lines

7) plotting can be with lines, marks, or points, optionally showing the alarm status for each point; when plotting is for means and standard deviation, error bars can be plotted; plotting can be value vs time, value vs value, or value vs sample number; axis scaling can be based on channel operating ranges, ranges of data values, or user-specified ranges; plots can be shown on individual grids or on a shared grid; for plotting on a shared grid, lines are automatically distinguished by color or dashed line pattern, depending on the plotting medium; plots can be on screen, to a PostScript printer, or to an Encapsulated PostScript file

8) the retrieved data can be exported to disk file; alarm status for each point can optionally be written; file can be in spreadsheet format or in archiver text file format; new export formats can easily be added

9) allows interactive selecting of channels to retrieve, interactive modifying of the list, and saving the list to a file for use in subsequent retrievals

10) can be operated in interactive mode or script mode

11) in both modes of operation, the archive control file can be used as the primary retrieval control file, to avoid the need to explicitly specify the name and type of the archive data file; in addition, the retriever supports editing the archive control file, primarily as a means for capturing annotations about the data.

References

[1] L.R. Dalesio, et al, "EPICS Architecture", Proceedings of the International Conference on Accelerators and Large Experimental Physics Control Systems, Tsukuba, Japan, Nov. 11-15, 1991.

[2] L.R. Dalesio, et al, "A Distributed Timing System For Synchronizing Control and Data Correlaton", Proceedings of the ICALEPCS, Tsukuba, Japan, Nov. 11-15, 1991.