DATA ARCHIVING SYSTEM IN KEKB ACCELERATORS CONTROL SYSTEM

T. T. Nakamura, N. Yamamoto and T. Katoh
KEK, Tsukuba, 305-0801, Japan

ABSTRACT

A data archiving system "KEKBLog" was developed in KEK for KEKB accelerators. KEKBLog consists of a simple archiving program, which is based on EPICS channel access protocol, and retrieving tools. Software structure of KEKBLog and hardware configuration are presented. KEKBLog has accumulated large amount of archive data and has served them for daily operation of KEKB accelerators since 1998, the beginning of the commissioning of KEKB. Many application programs that access to the archive data has been developed by the commissioning group. Current status of the usage of KEKBLog and recent progress are also reported.

INTRODUCTION

KEKB is an asymmetric electron-positron collider at 8 3.5 GeV/c, which is dedicated to B-meson physics. Its operation was started in December 1998. The KEKB accelerators control system has been constructed based on EPICS (Experimental Physics and Industrial Control System) tool kit [1]. EPICS provides core mechanism for the distributed control system. EPICS runtime database is running on a local control computer called IOC (Input/Output Controller). More than 100 VME computers are installed as IOC in the KEKB accelerators control system. Several workstations of 4 kinds of platform (PA-RISC/HP-UX, Alpha/OSF1, PC-AT/Linux and Macintosh/OSX) are also installed. Most of the higher level application programs run in these workstations. EPICS provides a network protocol called Channel Access (CA) to exchange data between computers.

Although EPICS tool kit includes data archiving system called Channel Archiver, we do not use it because it had not yet been released in 1998. Instead a data archiving system called "KEKBLog" is used in the KEKB accelerators control system. The core software of KEKBLog is a data archiving program called "kblog" and data retrieving program called "kblogrd", those were developed by N. Akasaka at KEK in the early days of the commissioning. Later kblogrd have been modified several times by T. T. Nakamura.

SOFTWARE STRUCTURE OF KEKBLOG

Data Archiving

The data archiving program, kblog, is an EPICS CA client and runs as background process. Function of the kblog is rather simple. The kblog reads a list of channel names from a configuration file at the start up. Then it tries to establish CA connection to the specified channels and starts monitoring them. Whenever the value of the channel is changed, the value with timestamp is recorded to the data file. Because of the simple function kblog process is robust and stable in operation.

In usual operation kblog process is generated by supervisory shell script. It generates the kblog process with expiry option of 30 days and also restricts its data file size less than about 2GB. When the kblog process is terminated by expiry date or by file size limit, the shell script automatically generates new kblog process.

Data Format

A kblog process produces a pair of files, a control file and a data file. The control file contains the information of the channels, a few miscellaneous parameters and index table. The data file contains a sequence of events. Each event consists of channel identifier, timestamp and value of the channel. Special entries called Milestone are also recorded in the data file. Milestone is a snapshot of values of all channels and is recorded periodically. Timestamp and file position of the Milestone are recorded in the index table. When the process is regenerated, new pair of files is created.

Data Retrieval

The data retrieval program, kblogrd, is a command line program. It is the only way to access the archive data. No other methods like API library are available. All the application programs which read archive data use kblogrd command. Many application programs read archive data. Most of them are developed by commissioning group and used for daily operation of the accelerators. Because of its simple data format, data retrieval is not efficient. Especially retrieving long term data takes much time to read data sequentially.

Recently "Catalog" files are provided for user convenience. They are ASCII files which describe channel names of all archive files since 2002. Command line search tool is also available. It shows list of archive files which contains given channel names. The Catalog files are updated manually when the configuration files of kblog are changed.

Since the kblog process runs in PC/Linux, archive data have the little-endian byte order. Original kblogrd was designed to run only in same byte order machine as kblog was running. For this reason, users of big-endian machines, HP-UX and Macintosh, were not able to use kblogrd command directly. They used to run kblogrd in little-endian machine using remote shell command. Recently modified version of kblogrd which run in big-endian machine was developed. It makes retrieval operation much simple.

HARDWARE CONFIGURATION OF KEKBLOG

Figure 1 shows the schematic view of the hardware configuration for KEKBLog. 2 Linux PCs are dedicated to data archiving. All of the kblog processes run in one of them. Another PC is a spare machine, which is also used for test purpose. Each PC has 80GB+80GB RAID1 hard disk storage, to which archive data are written. Later, the archive data are moved to NAS (Network Attached Storage). Data moving is manually done every 1 or 2 weeks. Currently 4 NAS are used for KEKBLog. They use IDE hard disks and are inexpensive. 3 of them are fully filled with the data since 2002. Newest one has 120GB 8 RAID + 250GB 8 RAID. Latest data are kept in it. These dedicated machines, 2 PCs and 4 NAS, are connected to the KEKB control network. The application programs running in general purpose workstations can read archive data from these machines.
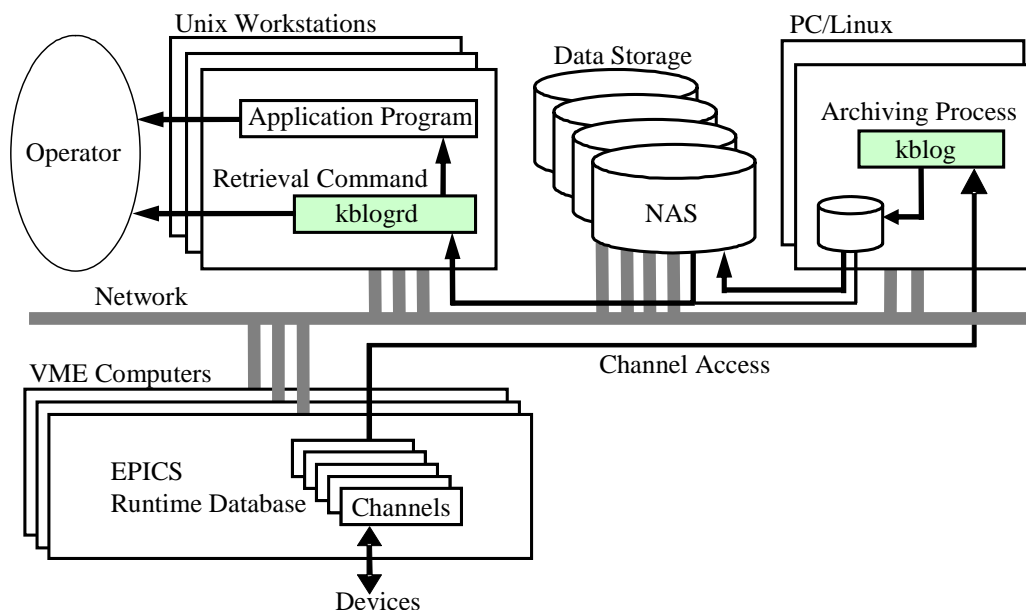

Figure 1: Schematic View of KEKBLog

Introducing dedicated hardware for data archiving makes maintenance easy and reduces downtime. Before 2001 kblog processes run in a general purpose server computer and sometimes had to stop archiving because of the machine shutdown for maintenance. Various problems and requests raised by other reasons than archiving often caused machine shutdown. In our case most of the problems came

from its large storage devices. After that, we have introduced separate hardware for archiving, mass storage and retrieval. And we also prepared dedicated machine for archiving and mass storage. Such a configuration reduces performance but is much easier to maintain.

DATA STORAGE

Table 1 shows current statistics of KEKBLog. 34 kblog processes run and totally about 3.5GB data are recorded everyday. Roughly speaking one year operation of KEKBLog produces 1TB archive data. Old data such as more than one year old are not necessary in most case. But in some rare case they are requested for reanalysis. In KEKBLog system all of the archive data are kept as long as possible. Recent progress of disk storage technology drastically reduces the cost of storage devices. Such a low cost storage device is suitable for our purpose. In KEKBLog system new NAS is added every year for these years.

Table 1: Current Statistics of KEKBLog

Category	Number of kblog processes (*1)	Number of monitoring channels (*1)	Amount of archive data (*2)
RF	2	7834	463 MB/day
Magnet	4	60596	191 MB/day
Vacuum	6	5902	413 MB/day
Beam Monitor	9	3318	1530 MB/day
Beam Transport Line	3	9323	565 MB/day
Injector Linac	4	1781	329 MB/day
Others	6	1372	111 MB/day
Total	34	90126	3602 MB/day

(*1) Numbers are determined on 22 Sep. 2005; (*2) Values are rough average over Feb.-Jun. 2005.

Data Backup

For data backup purpose we use HPSS (High Performance Storage System) in KEK Computing Research Center. It is a kind of hierarchy storage system, which consists of tape library and disks. Migration and staging between tapes and disks are automatically done. HPSS has 160TB tape library mainly for the physics experiment groups.

SUMMARY

KEKBLog has been developed and used in KEKB accelerators control system since the beginning of the commissioning. Archiving program kblog is simple and robust. But retrieving is not efficient because of its simple data format. Introducing dedicated computers for data archiving reduces maintenance complexity and allows stable operation. Inexpensive NAS is newly installed almost every year to store archive data. HPSS is used for data backup.

ACKNOWLEDGEMENTS

We wish to express our thanks to Dr. N. Akasaka, the original author of kblog and kblogrd. We also thank Dr. A. Morita for his heavy use of KEKBLog and many suggestions.

REFERENCES

[1] N. Yamamoto et al., "KEKB control system: the present and the future", PAC-99, New York, 29 Mar.-2 Apr. 1999.