

Algorithms and Data Structures for the EPICS Channel Archiver

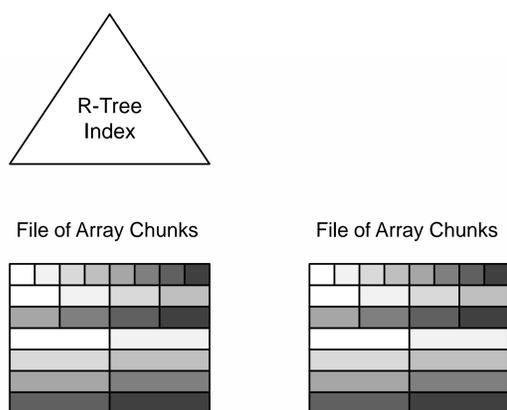
J. Rowland, M. T. Heron, S. J. Singleton, K. Vijayan, M. Leech,
Diamond Light Source, Oxfordshire, UK

Abstract

Diamond Light Source records 3GB of process data per day and has a 15TB archive on line using the EPICS Channel Archiver. This paper describes recent modifications to the software to improve performance and usability. The file-size limit on the R-Tree index has been removed, allowing all archived data to be searchable from one index. A decimation system works directly on compressed archives from a backup server and produces multi-rate reduced data with minimum and maximum values to support time-efficient summary reporting and range queries. The XMLRPC interface has been extended to provide binary data transfer to clients needing large amounts of raw data.

Channel Archiver

The EPICS R-Tree Channel Archiver uses a partitioned chunked array store with an R-Tree chunk index. SciDB and HDF5 (used for experimental data at Diamond) are also chunked array stores. The choice of data structure combined with storage properties determines performance.



I/O Utilization

Chunked array retrieval performance was limited by disk write IOPs on the old server as scattered writes competed with reads for controller interface bandwidth. The cache controller on the new server can retire many more write requests to battery-backed RAM. For comparison, the magnetic disk in the desktop machine without a write-back cache can only handle about 100 random IOPs/s, and will drop samples.

Hardware	Write%	Read%
New	3	75
Old	40	100
Desktop	100	100

Binary Retrieval

The EPICS Channel Archiver is an XMLRPC server; the XML codec is the limiting performance factor when retrieving large amounts of raw data. Using Base64 encoding to pack array samples into the XML payload improves performance while minimizing changes to existing clients.

Method	Space	Time
NFS	1x	10m20s
XMLRPC	20x	112m7s
XMLRPC Base64	1.3x	26m

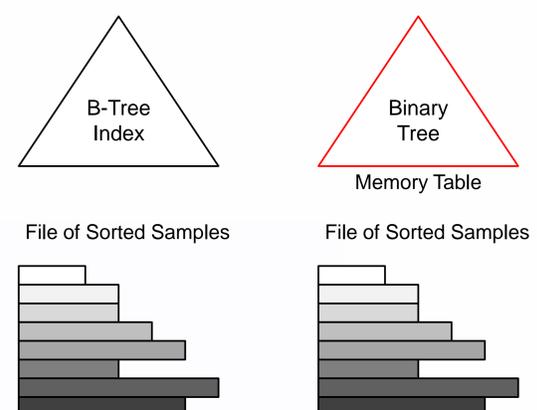
Benchmarks

The Diagnostics (DI) benchmark retrieves all beam positions for a period of one week, decimated for plotting. The Vacuum (VA) benchmark retrieves raw samples of all vacuum gauges for a period of one year. The VA benchmark was limited by XML codec performance before the introduction of the binary retrieval method.

Hardware	Test	Time
New	DI	30 s
Old	DI	3m30s
New	VA	8m30s
Old	VA	110m30s

Hypertable

Hypertable is a sorting key-value store with a software write-back cache. For EPICS channel archiving a suitable sort key is (channel, timestamp). Pairs are stored in memory in a balanced binary tree with a sequential log file for durability and periodically flushed to disk. This software caching allows good performance in the absence of a hardware write-back controller, and a similar method has been used at Diamond to prototype analysis and decimation tools.



Squashfs Compression

Archive data typically compresses by a factor of 4 due to the similarities between adjacent samples. The Linux squashfs file system offers transparent decompression and will increase storage capacity with no software effort.