

EPICS ARCHIVER APPLIANCE - INSTALLATION AND USE AT BESSY/HZB*

T. Birke[†], Helmholtz-Zentrum Berlin für Materialien und Energie GmbH, Berlin, Germany

Abstract

After 2 years of tests and development, the EPICS Archiver Appliance went into operation at HZB/BESSY in April 2018. After running for a year as an optional new archiver, the EPICS Archiver Appliance switched places with the old Channel Archiver and is now the central productive archiver in currently three installations at HZB.

To provide a smooth transition from the Channel Archiver to the EPICS Archiver Appliance for end users, some frontends like, e.g., the ArchiveViewer and other applications needed some modifications to be fully usable. New retrieval frontends are also provided and will replace the ArchiveViewer in the future. In addition, the versatile retrieval API rapidly improved the development of Python applications for analysis and optimization. Experiences with installation, configuration, maintenance and use of the EPICS Archiver Appliance will be shared in this paper.

INTRODUCTION

History

Since the very beginning of user operation at BESSY-II in 1998, the Channel Archiver was the newest technology to archive control system data in an EPICS [1] based control system environment. The Channel Archiver came with a few basic components:

ArchiveEngine The backend application to monitor a configured set of process variables and create archiver datafiles on a filesystem as well as an associated index file.

ArchiveDataServer an XMLRPC-server, that provides a web based API to let clients retrieve data from the archives.

ArchiveDaemon daemon to control starting and stopping a set of ArchiveEngines (see Fig. 1).

ArchiveDataTool, ArchiveManager, ArchiveExport, ... a set of commandline tools to manage archived data directly on the archive server.

ArchiveViewer an easy to use Java client application to search for archived PVs and retrieve, plot and correlate data graphically (see Fig. 2).

This setup has been a reliable workhorse for many years and users gladly accepted the ease of use of the ArchiveViewer to access data from anywhere in the world, started using JavaWebStart from the central BESSY/HZB Archiver Web page. In addition, generic tools

Engine	Port	Started	Status	Restart	Action
BESSY - Cryo	4804	08/19/2019 09:22:09	9999 channels connected	Every Monday at 02:00 Next stop 2019/08/08 02:00:00	Disable
BESSY - BPM-System	4803	08/19/2019 09:24:12	32243358 channels connected	Every Monday at 01:00 Next stop 2019/08/08 01:00:00	Disable
SRFGun-Test-Stand	4818	08/19/2019 09:31:12	216216 channels connected	Every Monday at 04:20 Next stop 2019/08/08 04:20:00	Disable
BESSY - Vacuum/Temps	4812	08/19/2019 09:31:16	20365048 channels connected	Every Monday at 03:20 Next stop 2019/08/08 03:20:00	Disable
BESSY - Timing	4811	08/19/2019 09:34:17	377077 channels connected	Every Monday at 02:00 Next stop 2019/08/08 02:00:00	Disable
HoBCat-Test-Stand	4814	08/19/2019 09:34:27	9991142 channels connected	Every Monday at 03:50 Next stop 2019/08/08 03:50:00	Disable
BESSY - XBFM	4813	08/19/2019 09:35:08	430439 channels connected	Every Monday at 03:30 Next stop 2019/08/08 03:30:00	Disable
BESSY-Beamline-Control-System	4801	08/19/2019 09:35:32	2513627009 channels connected	Every Monday at 01:30 Next stop 2019/08/08 01:30:00	Disable
BESSY - Feedback-Cameras	4806	08/19/2019 09:47:51	9191379 channels connected	Every Monday at 02:20 Next stop 2019/08/08 02:20:00	Disable
BESSY Control-System - Surveillance	4802	08/19/2019 09:50:10	21782274 channels connected	Every Monday at 01:40 Next stop 2019/08/08 01:40:00	Disable
BESSY - Beamcurrent-Measurement	4805	08/19/2019 09:55:11	14301489 channels connected	Every Monday at 02:10 Next stop 2019/08/08 02:10:00	Disable
BESSY - IDs	4808	08/19/2019 09:57:39	17852020 channels connected	Every Monday at 02:40 Next stop 2019/08/08 02:40:00	Disable
MLS - backend archiver	4817	08/19/2019 09:58:37	674910488 channels connected	Every Monday at 01:40 Next stop 2019/08/08 01:40:00	Disable
BESSY - HF	4807	08/19/2019 10:03:09	64148055 channels connected	Every Monday at 02:20 Next stop 2019/08/08 02:20:00	Disable
BESSY - Interlocks	4809	08/19/2019 10:17:20	33083008 channels connected	Every Monday at 02:50 Next stop 2019/08/08 02:50:00	Disable
BESSY - Power/Suitches	4810	08/19/2019 10:20:02	187914028 channels connected	Every Monday at 03:00 Next stop 2019/08/08 03:00:00	Disable
BESSY Insertion Device - Diagnostics	4815	08/19/2019 10:28:12	8090 channels connected	Every Monday at 04:00 Next stop 2019/08/08 04:00:00	Disable
BESSY Miscellaneous - (our Monochromator Control)	4816	08/19/2019 10:28:15	2461 channels connected	Every Monday at 04:10 Next stop 2019/08/08 04:10:00	Disable

Figure 1: ArchiveDaemon with 18 ArchiveEngines configured.

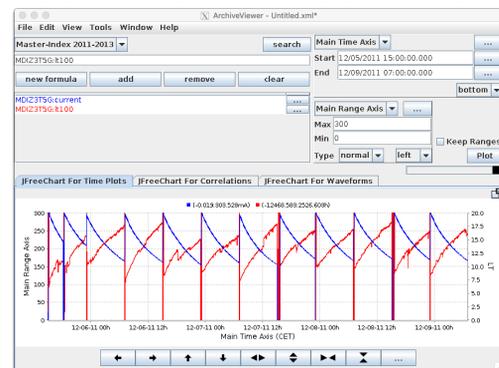


Figure 2: Original version of ArchiveViewer developed by S. Chevtsov especially for Channel Archiver.

have been developed to provide mostly pre-configured views (shift-handover, vacuum history...).

With the amount of data being archived growing over the years (see Fig. 3) managing the size-limited archive index files of 30 TB of data became a serious issue for administrators as well as for users.

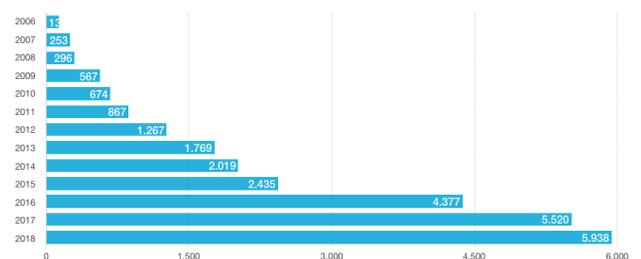


Figure 3: Growth of consumed storage by Channel Archiver data from 2006 until 2018 in GB/year.

* Work funded by BMBF and Land Berlin
[†] thomas.birke@helmholtz-berlin.de

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2019). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

EPICS Archiver Appliance

In 2015, M. Shankar (SLAC) released a work-in-progress version of the EPICS Archiver Appliance [2]. Even this early implementation solved many problems we ran into with the Channel Archiver. A first instance has been setup to archive a few high-volume PVs for a three months sliding window. It has been used since then mostly for performance and post-mortem analysis with data that couldn't be archived with the main Channel Archiver due to limited hardware resources.

From day one on, scientists and engineers very much appreciated the easy access to archived data from their analysis applications, that are exclusively written in python.

CURRENT INSTALLATION

After a test-period and a few bug-fixes, a new instance has been installed in 2017 to replace the Channel Archiver after a period of testing. While running in parallel with the Channel Archiver for 6 months, the most important end-user tool ArchiveViewer has been adapted (see Fig. 4) and fixed and finally the Archiver Appliance has become the main central control system data archiver at BESSY/HZB, Table 1.

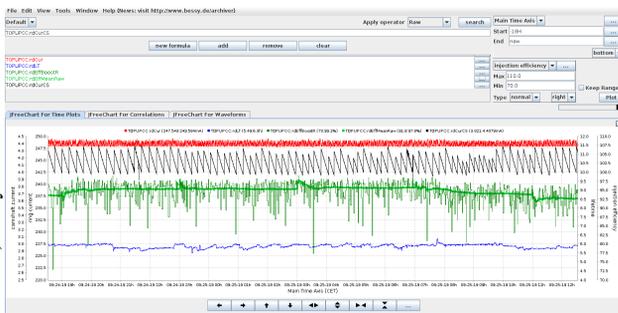


Figure 4: ArchiveViewer adapted by M. Shankar to also process EPICS Archiver Appliance Data.

Table 1: Existing Archiver Appliance Instances at HZB

Instance	# of PVs	GB/a	Comment
FASTZC	612	24000	high volume PVs kept for 3 months
BESSY	66658	5500	
bERLinPro	1408	50	rapidly increasing
MLS	13515	560	expected early 2020

The Channel Archiver is still running as a fallback archiver, since not all generic applications accessing archived data have been updated yet.

Meanwhile a third Archiver Appliance instance has been installed and set-up for the Berlin Energy Recovery Linac bERLinPro (see [3])

Hardware for a fourth instance at the Metrology Light Source (MLS) is already in house and will start archiving data beginning of 2020.

AUTOMATED CONFIGURATION

Configuration for the Channel Archiver is still maintained and/or generated in parts within the development environments of different subsystems. These parts are then combined to produce the Channel Archiver XML configuration files by a fully automated process.

The very same resulting XML files are now also used to generate a list of necessary configuration changes for each instance of the Archiver Appliance. These configuration changes are then applied automatically.

SYSTEM ENVIRONMENT

Server-Hardware

The current standard hardware platform for archive-servers is a system with 2 CPUs, 10 cores each and 256 GB RAM. Internal storage is provided using two fast NVME-SSD with a size of 2.4 TB organized as a RAID-1.

Storage

The EPICS Archiver Appliance features three levels of storage.

- **STS** - Short-Term-Store: A RAM-disk is used to store the latest hours of data.
- **MTS** - Medium-Term-Store: The latest month of data is stored in daily archives on an internal NVME SSD
- **LTS** - Long-Term-Store: SAN-storage - potentially endless currently stored in yearly archives.

Note: The currently chosen LTS granularity of yearly archive files imposes major problem on our backup system. About 60,000 datafiles grow on a daily basis - some only by a tiny bit - which forces the daily incremental backup to save these files every day. Towards the end of a year, this sums up to about 6 TB of backup data per day. Plans to try Hierarchical Storage Management (HSM) [4] with different partition granularities (e.g. PARTITION_MONTH for low data-rate PVs and PARTITION_DAY for high data-rate PVs) have just started testing on a separate test-instance.

EPICS Archiver Appliance

The EPICS Archiver Appliance consists of four services per instance that are running as Java-services behind a Tomcat application server each.

mgmt The management interface used to monitor and configure an archiver appliance instance. It provides also an API to retrieve information about the instance or add or delete PVs to be archived.

engine The engine is the actual archiving process. It connects to all configured PVs and writes received data to STS. The engine corresponds to the ArchiveEngine of the Channel Archiver.

retrieval The retrieval interface to the Archiver Appliance. It provides the API to retrieve archived data. The *retrieval* service corresponds to the ArchiveDataServer of the Channel Archiver.

etl ETL is an abbreviation for extract/transform/load. This describes the process, that performs transporting stored data from STS→MTS and from MTS→LTS, which both may involve e.g. steps to reduce the amount of data.

USER INTERFACE

ArchiveViewer

The well known ArchiveViewer, that was originally developed by S. Chevtsov for the Channel Archiver about 20 years ago, has been extended to also allow retrieval of data from the EPICS Archiver Appliance.

Any other further development of this application has been stopped several years ago and the code base can be considered frozen. The JavaWebStart mechanism that is used to run the ArchiveViewer without installation has been discontinued by Oracle® and is no longer available in latest Oracle®-Java releases. The OpenWebStart-project keeps development going, but since the ArchiveViewer is the only application making use of this mechanism at HZB, it is considered deprecated as well.

CSS/phoebus

The implementation of the DataBrowser in CSS/phoebus could likely be a successor to the ArchiveViewer. As a standalone application it has to be installed on every system to use it on. BESSY/HZB provides preconfigured Linux-, MacOS- and Windows-binaries for users to download and run on their desktop computers.

DataViewer

The simple DataViewer is capable of plotting waveforms over time in surface-plots. The cause of beam perturbation can be immediately visualized, as shown in Fig. 5.

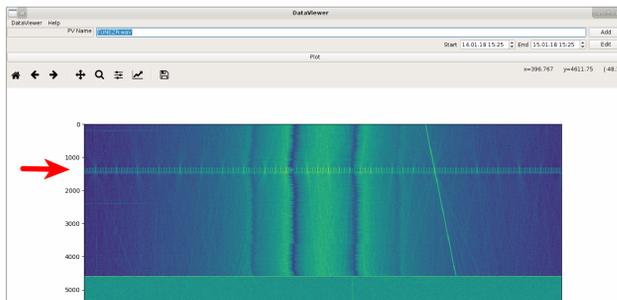


Figure 5: Simple DataViewer.

EPICS Archiver Appliance Viewer

The EPICS Archiver Appliance Viewer provides a useful web-based alternative to standalone programs (see Fig. 6).

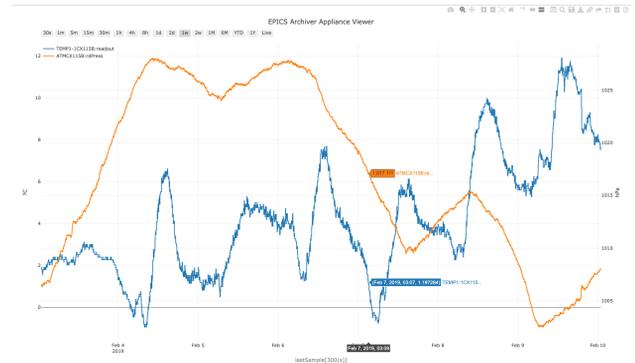


Figure 6: Web-based EPICS Archiver Appliance Viewer.

It provides most features needed for quick browsing through archived data. Nevertheless, it is not (yet) suitable as a full replacement for the CSS/DataBrowser or even the ArchiveViewer.

CGIexport

was originally developed for the Channel Archiver, this simple form-based web-interface provides easy access to data in several formats as well as images of plots via simple http-requests. With ~10000 requests per day, it is heavily used (typically periodically re-generated plots in web-pages) and hence the motivation to provide a fully compatible version that retrieves data from the Archiver Appliance is very high (see Figs. 7 and 8).

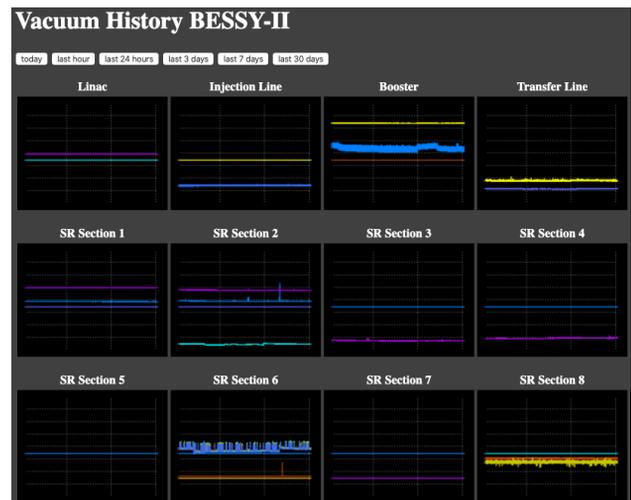


Figure 7: Vacuum history in a simple web view - made for the colleagues responsible for the vacuum system to get a quick overview of the development of the vacuum e.g. after a shutdown.

Archiver Appliance API

The retrieval API of the Archiver Appliance provides functions to retrieve data in various formats. Especially raw PB (Google Protocol Buffer) data is a very performant way to retrieve data.

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2019). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI.

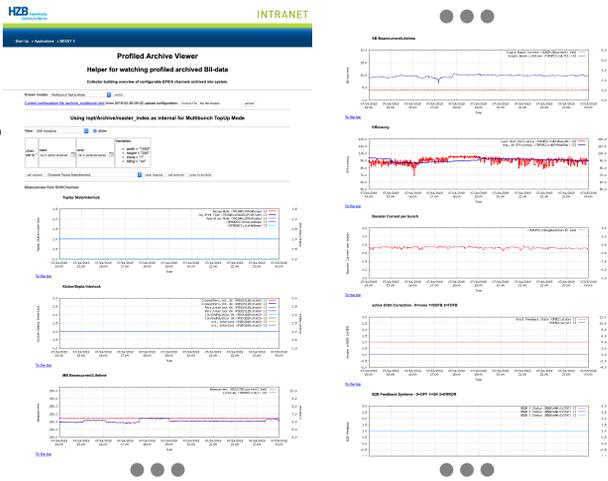


Figure 8: BESSY Shift-Handover Overview - an example of generically produced web-page using the form-based CGI as a backend to generate plot-images.

A small python layer has been developed, that automatically retrieves data in raw PB format and transforms this data into easy-to-use high performance pandas data structures, ready to be analyzed with powerful tools (Fig. 9).

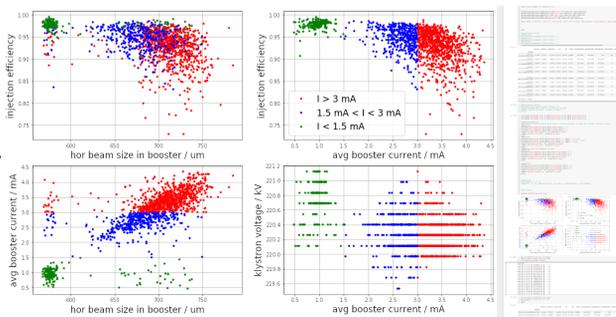


Figure 9: Data analysis performed on archiver data in jupyter notebooks. In this example, the injector performance is analyzed with different correlations.

The use of the python layer is as easy as:

1. Pick archiver instance to get data from,
2. Query and retrieve data by PV-list and time-range, which is automatically transformed from raw PB format into pandas data frames,
3. Process, analyze, plot/print data.

Besides analyzing machine performance or doing post mortem analysis, the archived data is also used by novel tasks like Machine Learning used in analysis and optimization toolsets [5] as training data.

LOCAL CHANGES

Some local changes have been made to the sources, especially to the ArchiveViewer, but a few minor fixes also to the Archiver Appliance itself. These changes still need to be

submitted back to the original authors, which will happen soon, once the source code concerned is cleaned.

ACTION POINTS

- Issues with the backup system will be solved, tests are already being performed to establish a new filesystem.
- There's still work to do adapting all client scripts and applications before the Channel Archiver can finally be abandoned and all queries can be performed using the Archiver Appliance without exception.
- Check local changes into central repository and create pull request.
- An alternative to ArchiveViewer needs to be provided.

CONCLUSION

- Help from Murali Shankar (SLAC) in understanding and fixing problems was very instructive and very much appreciated.
- Automated data-retrieval and -analysis with python application increased by a magnitude.
- Instant advantages were gladly accepted by users:
 - Response time of queries improved
 - No more split archives due to limited index size (no need to pick the right one).
- Maintenance effort has vastly decreased.

Overall, switching to the EPICS Archiver Appliance has proven to be a full success and a giant leap forward compared to Channel Archiver regarding administration aspects as well as using it as an information source.

ACKNOWLEDGEMENTS

Murali Shankar (SLAC), Dennis Engel, Frederik Hammerschmidt, Stine Heise, Markus Ries, Andreas Schällicke, Peter Stange, Luis Vera Ramirez and all the users reporting inconsistencies and problems in our setup as well as in the software.

REFERENCES

- [1] EPICS, <https://epics-controls.org/>
- [2] EPICS Archiver Appliance, Murali Shankar, SLAC, Stanford, CA, USA, https://slacmshankar.github.io/epicsarchiver_docs/,
- [3] T. Birke *et al.*, "Status of the Control System for the Energy Recovery Linac bERLinPro at HZB", presented at the 17th Int. Conf. on Accelerator and Large Experimental Control Systems (ICALEPCS'19), New York, NY, USA, Oct. 2019, paper FRAPP06.
- [4] Hierarchical storage Management, https://en.wikipedia.org/wiki/Hierarchical_storage_management

- [5] L. Vera Ramirez, T. Mertens, R. Mueller, J. Viefhaus, and G. Hartmann, "Adding Machine Learning to the Analysis and Optimization Toolsets at the Light Source BESSY II", presented at the 17th Int. Conf. on Accelerator and Large Experimental Control Systems (ICALEPCS'19), New York, NY, USA, Oct. 2019, paper TUCPL01, this conference.