

Evaluation of a new visualization and analytics solution for slow control data for large scale experiments

Laura Sargsyan^{1,*}, and Filipe Martins²

¹A. Alikhanyan National Science Laboratory, Armenia

²Laboratory for Instrumentation and Experimental Particle Physics, Portugal

Abstract. Large experiments in high energy physics require efficient and scalable monitoring solutions to digest data of the detector control system. Plotting multiple graphs in the slow control system and extracting historical data for long time periods are resource intensive tasks. The proposed solution leverages the new virtualization, data analytics and visualization technologies such as InfluxDB time-series database for faster access large scale data, Grafana to visualize time-series data and an OpenShift container platform to automate build, deployment, and management of application. The monitoring service runs separately from the control system thus reduces a workload on the control system computing resources. As an example, a test version of the new monitoring was applied to the ATLAS Tile Calorimeter using the CERN Cloud Process as a Service platform. Many dashboards in Grafana have been created to monitor and analyse behaviour of the High Voltage distribution system. They visualize not only values measured by the control system, but also run information and analytics data (difference, deviation, etc.). The new monitoring with a feature-rich visualization, filtering possibilities and analytics tools allows to extend detector control and monitoring capabilities and can help experts working on large scale experiments.

1. Introduction

Monitoring of the control system data of large scale experiments has always been a challenging task. Control systems serving the accelerator, detector complex, and the supporting technical infrastructure generate very large amounts of data. With this overwhelming amount of data comes the challenge of analysing it in a way that brings big value. Analysis of this data requires efficient, scalable and flexible technologies.

Higher data frequency with the FPGA equipped devices, more comprehensive information from the device statuses, new demands to the data mining requires new approaches. The current DCS Data Viewer (DDV) [1] is used to access and represent DCS historical data and does not include analytics tools. Envisioning future challenges, we propose different approaches and technology solutions to improve the level of the service

* Corresponding author: Laura.Sargsyan@cern.ch

provided. There is a need to be updated with the emerging big data technologies. The proposed solution leverages new virtualization, data analytics, and visualization solutions.

The visualization and analytics tools utilize the OpenShift Platform running on CERN Cloud [2]. OpenShift leverages Docker [3] containers and Kubernetes [4] orchestration to provide a Platform-as-a-service (PAAS) solution oriented for web applications. The web application hosting service allows building instances using different application templates, as well as deploying custom applications.

The prototype of the new visualisation and analytics system was developed to analyse the High Voltage (HV) distribution to the PhotoMultiplier Tubes (PMTs) of the ATLAS [5] Tile Calorimeter (TileCal) Demonstrator and Extended [6] barrel modules. The Demonstrator prototype was built to evaluate and qualify the new read-out electronics for the ATLAS Phase-II upgrade.

In this contribution the integration of InfluxDB [7], the times-series database (DB) is described along the experience of using Next Generation Archiver (NGA) [8] with Influx Backend.

Access to the Detector Control System (DCS) data and visualization done using an open-source metric analytics and visualization suite – Grafana [9].

The paper presents the architecture, implementation details, features, functionalities of a new visualization and analytics system, as well as problems encountered, and plans.

2. New DCS monitoring and analytics system

2.1 DCS monitoring data flow

The DCS monitoring data flow is shown in Figure 1. It consists of the following components: Data Source, Transport, Storage, and a Data Access.

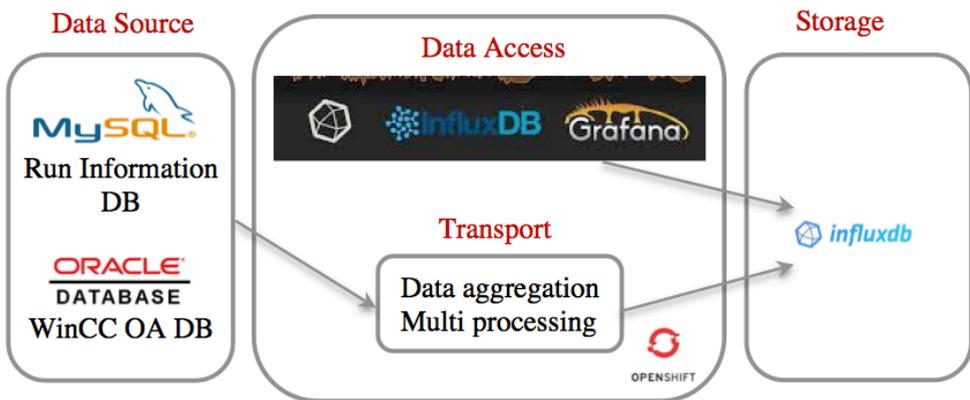


Fig. 1. DCS monitoring data flow.

- Data Source: the data collected from the DCS ORACLE archive and Run Information MySQL databases.

- Transport: data collectors retrieve information from the Data Source. Collectors are running using Docker methodology. Docker allows separate applications from the infrastructure, so the software can deliver quickly. More information in section 2.2.
- Storage: the data collection is stored in InfluxDB time-series database. InfluxDB can handle hundreds of thousands of values per second and offers different retention policies according to the user needs.
- Data Access: data visualization done in Grafana - the open source web based platform for monitoring and analytics.

2.2 Monitoring system architecture on CERN OpenShift Cloud Computing Platform

Over the past of years, CERN has adopted cloud technology that allows users' access over the Internet to shared pools of resources. Cloud deployments offer support for load balancing, circuit breaking and rolling updates in a near-transparent way. The CERN OpenShift PAAS Cloud Computing model is a Kubernetes container platform with full-stack automated operations to deploy and manage multi-container applications at scale.

The monitoring system architecture on the CERN Cloud is illustrated in Figure 2. The OpenShift project includes several Docker containers for different use-cases. The application templates provide in the catalog allow installing the needed web application without hassle.

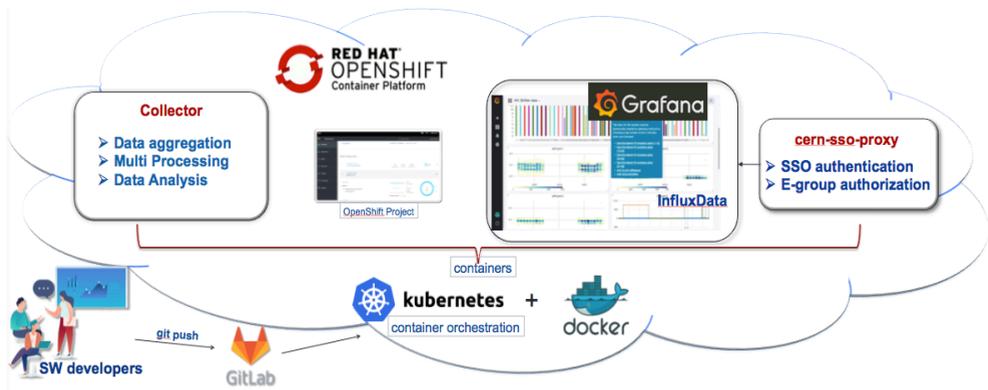


Fig. 2. Monitoring system architecture.

Thanks to image provisioning automation, Grafana and cern-sso-proxy applications were installed and configured to provide access to CERN authentication services in a very short time. Apart from the application templates, OpenShift offers a source-to-image build system to automatically create container images from the custom applications' source code. The custom python applications, data collectors, were developed to transfer physics Run information (start time, end time, Run type), DCS HV data, and also calculate analytics data (difference, deviation, etc.) to the InfluxDB. The docker container had been created and configured. The build and deployment of the collectors were done in minutes.

GitLab's built-in Continuous Integration/Continuous Delivery (CI/CD) [10] is integrated with OpenShift deployment. This gives a possibility to deploy a modified application from

GitLab to the OpenShift container cluster, e.g. when the code in the GitLab repository master branch is updated, the new build and deployment are sequentially triggered.

OpenShift Container Platform accelerates the application development, optimizes the use of the infrastructure, and provide distributed, advanced and scalable services based on the Docker container technology and orchestrated by Kubernetes.

3. Grafana Visualization

The new DCS monitoring solution was created to monitor and analyze DCS data to improve operations. Plotting multiple graphs in the control system, extracting historical data for the long period using a trending widget in WinCC Open Architecture (OA) [11] is a resource-intensive task. DCS graphical user interface could become unresponsive. In the case of the control system, it is not acceptable. Visualization in the presented system performed using a tool that runs independently from the control system, so loading on the control system computing resources is decreased. Grafana was chosen as a visualization platform. It is a feature-rich open-source solution for data visualization, monitoring, and analysis.

Currently many dashboards were created to study the High Voltage applied to the PMTs for the new hardware that was installed in a test-beam hall on CERN's Preveessin site. Dashboards display DCS and analytics data using different type of Grafana panels: Graph, Table, Text, Heatmap, Histogram, etc. Dashboard contains several individual panels with different functionalities.

High Voltage evolution dashboard is presented in Figure 3. It displays integrated monitoring utilizing data from two sources - DCS Oracle DB and MySQL Run information DB. The dashboard shows metrics for different modules - Extended Barrel and Demonstrator – next to others. This was done to compare the behavior of the new and old hardware in one view. Users can choose filters (1): LedRun and/or EIRun. The result: Run annotations (vertical lines) will show number, type, end, and start time of the Run(s) for the defined time interval (2). The legend (3) shows time series measurements. References in (4) provide direct access to other dashboards.

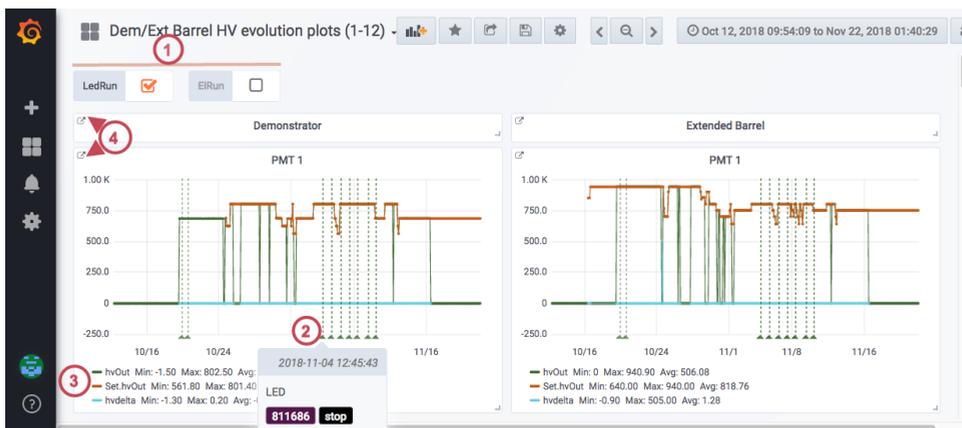


Fig. 3. Demonstrator vs. Extended Barrel dashboard.

Grafana can query and efficiently process historical data in large time window using auto-scaling. Scaling factor in query changes synchronously with time range and interval changes. For a short time range, e.g. for 24 hours, the 5 min resolution is used. For longer

time range resolution increases accordingly. Thanks to this feature, users can navigate without the risk of hanging up their browsers.

In some cases, users need more detailed information. The “Shifter view” dashboard was created for this purpose. It shows all data, but previously user gets the total number of values and then should decide whether navigate further or decrease the time window to reduce the number of retrieved values (Figure 4).

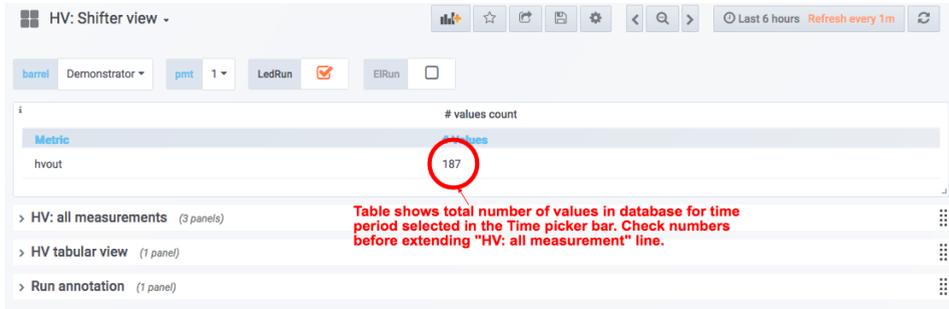


Figure 4. “HV: Shifter view” dashboard.

Grafana, by default, stores configuration data (dashboards, users, etc) in the SQLite database. During exploitation, the issue was observed. Grafana with SQLite backend becomes very slow when it has many concurrent requests. Moving configuration data from SQLite to the MySQL database resolves the problem.

In Figure 5 “Run table with HV reference” panel is presented. It provides access to the HV data with different views for the selected Run. This feature was requested by physics analysis users to qualify the hardware during physics Runs and validate data.

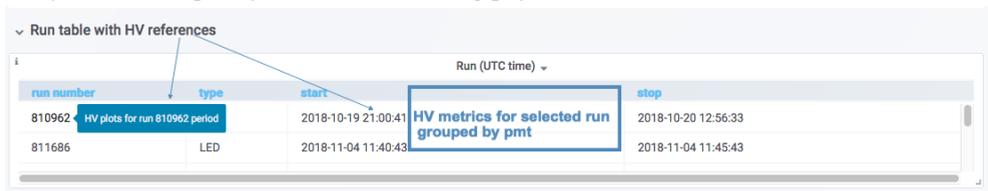


Figure 5. “Run table with HV reference” panel.

High Voltage values are set using DCS system. The applied HV (hvOut) should be stable with spread (hvdelta) smaller than 0.5 V. Figure 6 shows hvOut and hvdelta values grouped by PMT. The graph helps to spot wrongly working PMT channels.



Fig. 6. HV and HV delta values grouped by PMT.

Many different dashboards were developed to meet the requirements of diverse categories of users - engineers, researchers, shifters, and operations team.

4. Next Generation Archiver

The Next Generation Archiver (NGA) project is developed in collaboration between CERN and Siemens/ETM for WinCC OA systems. The main motivations are:

- Provide scalable archiving solution required for next upgrades of the LHC and its experiments beyond 2020.
- Enable to use more advanced analytics tools.
- Enable the usage of NOSQL (Not Only SQL) data store and processing technologies.

The NGA component with the Influx backend was installed and validated on a shadow WinCC OA project in the laboratory. Values from more than 1000 data point elements (DPEs) started to archive in InfluxDB. In order to decrease the insertion data rate the smoothing of archived data was implemented. The average rate of archived data is about 2 MB per day, which is almost matched to the data rate of the production system. During validation, no issues were observed.

Figure 7 shows temperature probes inside drawer using data archived by NGA. With the NGA, efforts to transfer DCS data from ORACLE to InfluxDB were reduced.



Fig. 7. Temperature probes inside drawer.

5. Conclusions and Outlook

Anticipating future challenges we propose a powerful tool to digest DCS data. The analysis of this data is crucial for evaluating and improving the performance, and efficiency of the control system. The prototype of the new visualization and analytic platform was delivered for the test environment. In the future, we plan to continue the validation of NGA with Influx backend on the production environment. Improving and enlarging of analytics solutions using machine-learning tools is also foreseen.

The proposed visualization solution enhances the graphical user interface with modern web based trends and graphs. It runs separately from the control system, thus reducing the workload on computing resources where the control system is launched. Integrated monitoring utilizes data from different sources and gives a broader view to the experts, simplifying work and speeding up issue tracking. The new archiving with Influx backend allows handling high write and query loads. Descriptive monitoring with analytics

capabilities opens new data exploitation approaches. Usage of the Cloud Computing simplifies and automates the service provisioning. The new monitoring can be used as a future visualization extension of the control system, saving time for engineers, researchers, and operation experts.

5. References

1. C. Tsarouchas, S. Schlenker, G. Dimitrov and G. Jahn, *Dcs Data Viewer, an Application that Accesses ATLAS DCS Historical Data*, J. Phys.: Conf. Ser, 513 (2014)
2. A. Lossent and A. Peon, *PaaS for web applications with OpenShift Origin*, J. Phys.: Conf. Ser, 898 (2017)
3. Docker Inc, *What is Docker?* [Online]: <https://www.docker.com/what-docker>
4. Kubernetes, *Kubernetes - Production-Grade Container Orchestration*, [Online]: <https://kubernetes.io/>
5. ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST 3 S08003 (2008)
6. ATLAS Collaboration, *Technical Design Report for the Phase-II Upgrade of the ATLAS TileCalorimeter*, Tech. Rep. CERN-LHCC-2017-019; ATLAS-TDR-028 (2017)
7. InfluxDB [Online]: <https://www.influxdata.com>
8. P. Golonka, M. Gonzalez-Berges, J. Guzik, and R. Kulaga, *Future archiver for CERN scada systems*, C17-10-08, p.THPHA037 (2018)
9. Grafana [Online]: <https://grafana.com>
10. GitLab, *Continuous Integration is built-in to GitLab* [Online]: <https://about.gitlab.com/product/continuous-integration/>
11. SIMATIC WinCC Open Architecture Portal [Online]: <https://www.winccoa.com/>