Migration to WebDAV in Belle II Experiment

Silvio Pardi[1], *Takanori* Hara[2], *Michel* Hernandez Villanueva[3], *Hideki* Miyake[2], *Cédric* Serfon[4], *Hiroaki* Ono[5], and *Ikuo* Ueda[2]

[1]INFN-Napoli Unit – Italy
[2]High Energy Accelerator Research Organization (KEK), Tsukuba 305-0801, Japan
[3]Deutsches Elektronen-Synchrotron (DESY), Notkestraße 85, 22607 Hamburg, Germany
[4]Brookhaven National Laboratory, Upton, NY, USA
[5]Nippon Dental University, 1-8 Hamaura-cho, Niigata 951-8580, Japan

Abstract. The usage of WebDAV protocol has become more and more popular within the physics experiments using grid middleware in the last decade, and today it represents a valid alternative to the GridFTP currently supported at best-effort level after the retirement of Globus Toolkit. Belle II experiment established the adoption of WebDAV protocol as the main protocol for data access and third-party-copy transfers, without relying on Storage Resource Manager interface (SRM). The migration process, carried on with continuous and gradual steps, has required a large effort to guarantee a smooth transition maintaining the production infrastructure fully operational. In this contribution we show the transition process, the tool of support developed to monitor step by step the status of third-party-copy support with WebDAV protocol by storages of the collaboration tested in both case pull and push, the strategy adopted to configure DIRAC and the solutions put in place for the corner cases. Finally, we will present some statistics of utilization and we will analyse the achieved results.

1 Introduction

Belle II is an international experiment with the goal of conducting a scientific program on Flavor Physics using data collected at the SuperKEKB e+e- collider located at the KEK laboratory in Tsukuba, Japan [1]. The primary objective of this experiment is to search for signals of new physics beyond the Standard Model by amassing approximately 50 times more data (~50ab-1) than its predecessor, Belle, which operated at the KEKB accelerator. At the peak luminosity, the experiment is expected to produce detector data on the order of O(10 PB) per year.

To effectively analyse this substantial volume of data, Belle II takes advantage of a distributed computing infrastructure [2,3], to perform several tasks such as RAW data processing, Monte Carlo production, physics analysis, data storage, and data archiving. Throughout its operations, the experiment extensively utilizes technologies adopted by the Worldwide LHC Computing Grid (WLCG) [4] technologies. Notably, the SRM [5] (Storage Resource Manager) protocol and GridFTP [6] have served as foundational components for data access and transfer in previous years.

However, due to the retirement of the community-supported portion of the Globus Toolkit, which currently is receiving best-effort support, an urgent necessity has arisen to

transition to alternative technologies. WebDAV [7] and xrootd [8] have emerged as two primary alternatives under consideration.

Belle II experiment decide to adopt WebDAV as the main protocol for data access and third-party transfers. The migration campaign started early 2022, with a gradual adoption of WebDAV by different storages. To govern this process a dedicated monitoring system has been implemented and used as troubleshooting tool. The migration process required a large effort to ensure a smooth transition while keeping the infrastructure operational. In this paper we will show all steps done to face the migration and the results obtained in terms of performance and reliability and wide adoption of WebDAV in the Belle II computing infrastructure,

The rest of the paper is organized as follows: In section 2, we provide an overview of the motivation and the migration process. In section 3, we introduce the monitoring system developed as a tool to systematically check third-party-copy storage to storage via WebDAV. In section 4, we analyse the performance of WebDAV using the DIRAC tools. Finally, in section 5, we summarize our work and draw conclusions.

2 WebDAV for Data Access and Third-Party-Copy transfer

Belle II Computing Infrastructure uses DIRAC [9] as Workload Management system framework, with a set of extensions included in the BelleDIRAC [10] release. The data management system is Rucio [11] which has been integrated in the Belle II computing system, and interfaced to DIRAC [12,13]. Rucio is then using FTS service for third party copy.

In the DIRAC configuration, for each storage we defined supported protocols through a set of variables that identifies the hostname that serves the specific protocol, the port, the base path, the protocols name and other parameters to properly identify the endpoint. Then through the variable *AccessProtocols* and *WriteProtocols* we respectively identify the protocol used for data access and for writing.

All the storage endpoint definitions in our DIRAC configuration are synchronized to the Rucio configuration by a system component (DIRAC Agent). To switch from SRM+gsiftp to davs (abbreviation of WebDAV) we needed to reconfigure all the storage elements, by defining the new davs protocol and setting it as first protocol in the list.

Fig. 1. Example of Storage configuration in production DIRAC installation for Belle II

3 A monitoring tools

To monitor the progress of WebDAV support across the different Belle II storage systems for third-party copying, a testing engine has been developed. The system sends a set of file transfer jobs via the production FTS service. For each pair of endpoints, the system sends two jobs, one using the push mode and the other the pull mode. A full set of tests is run four times per day. All results are published via the web, showing a matrix (figure 2) in which, a green square indicates that both push and pull transfers were successful, a yellow square indicates that only one direction worked properly (push or pull), while a red square represents the case in which none of the tests have been completed successfully. For each transfer is sufficient that one of the two modes works, so having yellow dominating for a specific storage in the plot does not represent a problem for protocol switching.

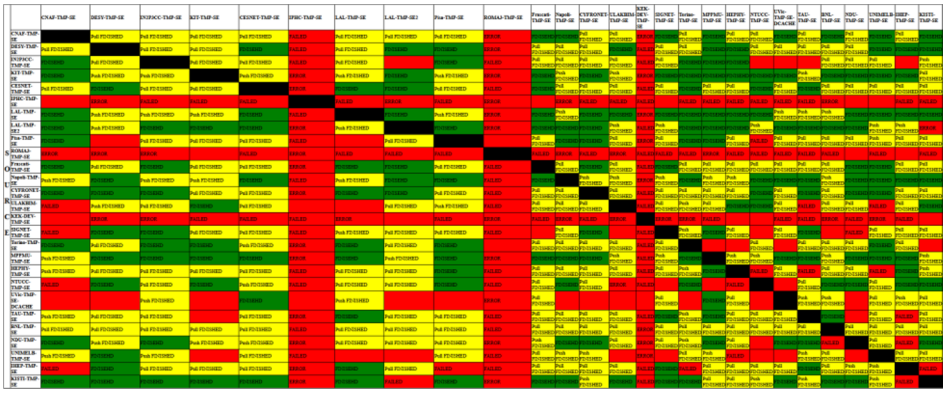

Fig. 2. Main matrix of the Third-Party-Copy monitoring dashboard in which each square represent the status of a FTS transfer job. In particular this picture show the status of transfer early when we started the migration process to WebDAV

A schema of the two modality pull and push mode is represented in figure 3 in which we see the third party copy client contacts the destination site and issues a request that the destination downloads (pulls) the data from the source over HTTP, while in case of push the client contact the source serving a Destination header.

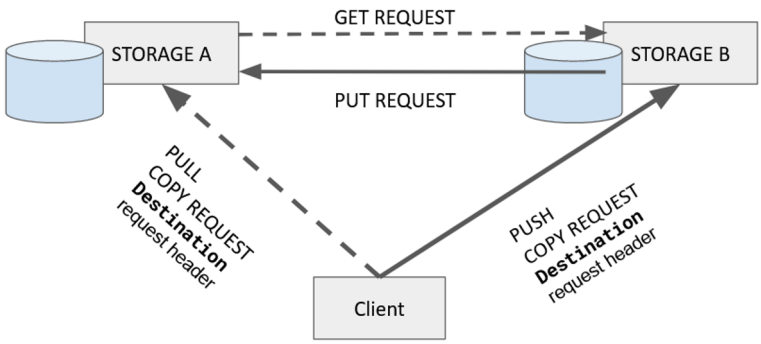

Fig. 3. In this figure pull mode is represented with dashed line whiles push mode with continuous lines

Together with the described matrix, the monitoring tools also provides two additional tables (Fig.4), one reporting the results of all the transfers ran in push mode and one for the one ran in pull mode. Such tables report again with green, yellow and red the status of the transfers, in addition each square, representing a single FTS job, provides the duration in second in case of successfully transfer, or the error message eventually reported by FTS together with a direct link to the corresponding job, in order to make smoother the troubleshooting of the specific issue.

Fig. 4. Detail of the pull matrix of the Third-Party-Copy monitoring dashboard.

In figure 5, it is observed the evolution of the third-party-copy transfer matrix since 2022. We began with a highly heterogeneous situation in which a large portion of the transfers failed, mostly due to storage misconfigurations or firewall restrictions. By meticulously addressing each error in the matrix, it has gradually improved. In the meanwhile additional storage resources were added. Since April 2023 the matrix is stabilized with only sporadic failures. Notice that one of the storages works only in pull mode because the local policy, that does not represent an issue as reported.

Fig. 5. The picture on the left shows the testing matrix as of Aprile 2023, the picture in the right shows the matrix one year after. The comparison proofs the large improvement obtained in term of stable WebDAV adoption by the sites.

4 Performance analysis

Since the start of the migration campaign in 2022, the adoption of WebDAV has steadily increased. Storage configurations have been adjusted to prioritize 'davs' as the primary protocol for data access and third-party-copy, while 'SRM,' 'gsiftp,' and 'root' are used as

backup protocols for most of the storage systems. The only exception is the tape system, which still operates through 'SRM.'

The graph in figure 6 displays the number of successful transfers by protocol since January 2022, between the computing resources (represented by the Worker Nodes of the various clusters serving the collaboration) and the Storage Elements. From the picture, it is evident that the number of SRM transfers has significantly decreased.

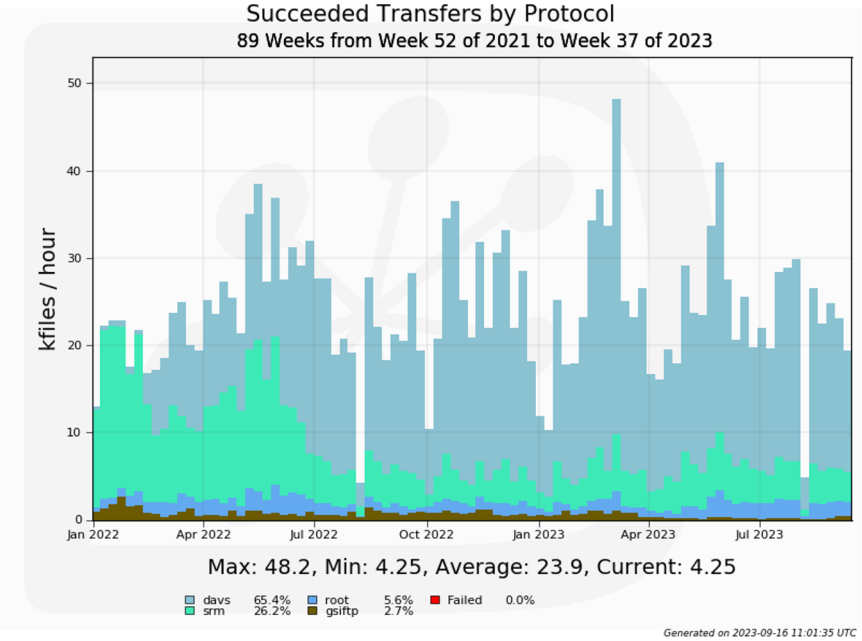

Fig. 6. Successful transfer Worker Node vs Storage Elements by protocol

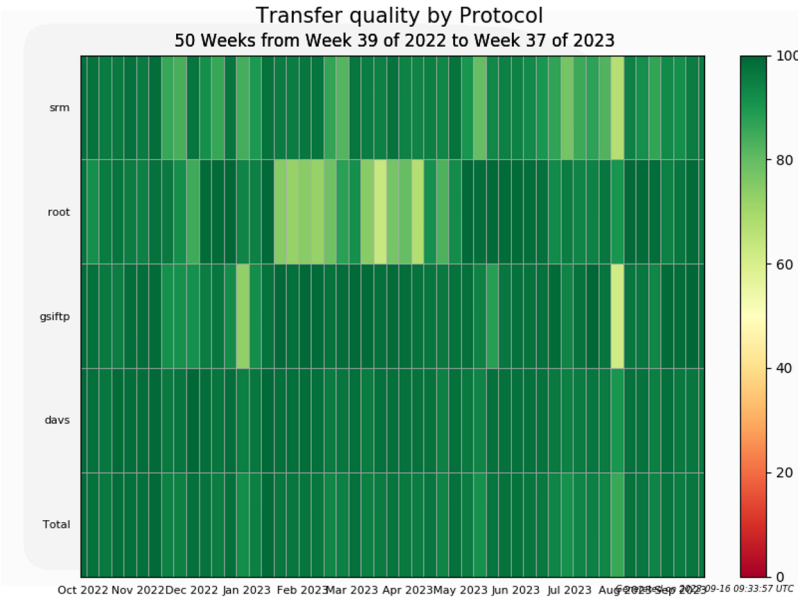

Fig. 7. Transfer quality by protocol graph shows the high level of efficiency of WebDAV in the last year of operation.

After 1 year of operation, WebDAV has demonstrated a high level of reliability. The matrix in figure 7 obtained using the DIRAC accounting tools, compare the efficiency by protocol of the transfers done via SRM, root, gsiftp and davs.

5 Conclusions

Belle II has invested a large effort in the comprehensive adoption of WebDAV as the primary protocol for data access and transfer, in substitution of SRM+GridFTP. The transition has successfully been accomplished for disk-based storages. The remaining crucial step entails migrating the TAPE system to the WebDAV protocol and embracing the use of REST APIs for tape staging.

This strategic shift towards WebDAV represents a significant milestone in enhancing our data management capabilities. This transition enhances the efficiency of data access and transfer and ensures compatibility with contemporary data management standards, ultimately contributing to the continued success of the Belle II project.

References

1. T. Abe, et al. [Belle-II Collaboration], arXiv: 1011.0352 [physics.ins-det]
2. S. Pardi et al. “Computing at Belle II” Nucl.Part.Phys.Proc. 273-275 (2016) 950-956 DOI: 10.1016/j.nuclphysbps.2015.09.148
3. T. Hara on behalf of the Belle II computing group 2015 J. Phys.: Conf. Ser. 664 012002 DOI 10.1088/1742-6596/664/1/012002
4. I. Bird “Computing for the large hadron collider Annual Review of Nuclear and Particle Science” 2011 61 1-517 https://doi.org/10.1146/annurev-nucl-102010-130059
5. A. Shoshani et al “Storage Resource Managers: Middleware Components for Grid Storage”, In proceedings of Mass Storage Systems conference, Maryland USA 200
6. W. Allcock W, Bresnahan J, Kettimuthu R, Link M, Dumitrescu C, Raicu I and Foster I 2005 The globus striped gridftp framework and server - Washington, DC, USA: IEEE Computer Society - pp 54– ISBN 1-59593-061-2
7. The IETF Trust 2007 HTTP Extensions for Web Distributed Authoring and Versioning (WebDAV) RFC 4918
8. A. Dorigo et al “A highly scalable architecture for data access in the root environment” - Stevens Point, Wisconsin, USA: World Scientific and Engineering Academy and Society (WSEAS) pp 46:1–46:6 ISBN 960-8457-11-4
9. F. Stagni et al “The DIRAC interware: current, upcoming and planned capabilities and technologies“, EPJ Web Conf. 245 03035 (2020). doi: 10.1051/epjconf/202024503035
10. H. Miyake et al., Belle II production system, J. Phys. Conf. Ser. 664, no. 5, 052028 (2015)
11. M. Barisits et al., Rucio: Scientific Data Management, Comput Softw Big Sci 3 11 (2019)
12. C. Serfon et al. ”The Rucio File Catalog in DIRAC implemented for Belle II" EPJ Web Conf. 02026 (2021). DOI 10.1051/epjconf/202125102026
13. C. Serfon et al. “Integration of Rucio in Belle II" EPJ Web Conf. 02057 (2021). DOI 10.1051/epjconf/202125102057