A performance study of WebDav access to storages within the Belle II collaboration

S Pardi¹ and G Russo¹,²
¹INFN-Napoli Unit – Italy
²University of Napoli Federico II – Italy
E-mail: spardi@na.infn.it

Abstract. WebDav and HTTP are becoming popular protocols for data access in the High Energy Physics community. The most used Grid and Cloud storage solutions provide such kind of interfaces, in this scenario tuning and performance evaluation became crucial aspects to promote the adoption of these protocols within the Belle II community. In this work, we present the results of a large-scale test activity, made with the goal to evaluate performances and reliability of the WebDav protocol, and study a possible adoption for the user analysis. More specifically, we considered a pilot infrastructure composed by a set of storage elements configured with the WebDav interface, hosted at the Belle II sites. The performance tests include a comparison with xrootd and gridftp. As reference tests we used a set of analysis jobs running under the Belle II software framework, accessing the input data with the ROOT I/O library, in order to simulate as much as possible a realistic user activity. The final analysis shows the possibility to achieve promising performances with WebDav on different storage systems, and gives an interesting feedback, for Belle II community and for other high energy physics experiments.

1. Introduction

High Energy Physics experiments, rely on a well-defined set of protocols for data access, data movement, and to support all aspects of data analysis. Such protocols have demonstrated to work properly during the last decade, enabling scientists to move and manage data across hundred sites distributed worldwide. Among them gridftp, xrootd, and SRM are certainly the most popular and everywhere implemented protocols, and can be considered de-facto standards for HEP. However in the last years some limits are emerged, that include technical issues and the long-term economic sustainability.

The exigence of support more standard protocols is became more and more crucial with the growth of users communities interested in porting their applications on the large e-infrastructures, built under leading role of HEP. Moreover resources to maintain and develop proprietary protocols like SRM or xrootd, specific for physics experiments, is becoming a critical aspect in regime of increasing needs and flat-budget. The possibility to use standard protocols with large supporting communities like HTTP and WebDav could help to reduce the pieces of software to be maintained, and mitigate the economic issues.

In the last years, we have assisted at a large diffusion of new paradigms for storage provisioning offering the possibility to use cheap online storage areas. In most case, such storages provide natively HTTP or
WebDav interfaces, together with cloud specific protocols as S3. These boundary conditions have contributed to promote the discussion within the HEP community to understand how experiments could take advantage from these resources, and in which way is possible to integrate standard protocols in the experiments’ computing models.

Belle II experiment has early started to investigate this topic with a specific R&D activity, in order to be ready to work in a more flexible environment in the next future and to contribute at the international discussion. Moreover the current Belle II production infrastructure include an increasing number of cloud resources which run MC production jobs, and that could even offer storage services.

In this paper we present a performance analysis work, done with the goal of evaluate the state of the art of HTTP/WebDav support provided by different storage technologies. To do that we have created a testbed using Belle II resources currently in production, and then we performed a set of benchmark to compare HTTP/WebDav protocols vs xrootd and gridftp.

The collected results have been analysed and presented in this work with the goals to proof that Belle II software can work in a full HTTP ecosystem, and give a snapshot of the current achievable performance.

The rest of the paper is organized as follows: in Section 2 we will summarize the context and the most relevant activities currently ongoing. In section 3 and 4 we describe the testbed infrastructure, and the steps needed to integrate the HTTP/WebDav protocols in belle II software. Then we summarize the test plan used to analyse protocol performance. In section 6 we analyse the results of the performed tests. Finally, we summarize the obtained results, and we give some hints for further investigation.

2. Context and related works

The usage of HTTP and WebDav as data access and storage management protocols, is largely discussed in High Energy Physics community. WLCG has created an ad-hoc taskforce [1] with the mandate to understand the requirements of the major experiments, and which solutions should adopted to accommodate the needs. This taskforce is evolved in a stable working group for Data Management. The topic is part of the discussion that goes with the name of “Protocol Zoo” [2] that aims at reduce costs for data management, and rationalize the number of protocols used in the HEP communities.

In the last years several efforts have been made to implement a full HTTP/WebDav ecosystem, that includes the integration of these protocols by the major storage systems such DPM, dCache, STORM, and the creation of libraries for the direct usage of HTTP URLs in the ROOT framework [3][4].

This work contributes at the discussion by studying how to integrate such protocols in the Belle II Computing model [5] and by analysing performance on the base of a set of representative tests, performed on a realistic testbed built over the Belle II production infrastructure.

3. The storage infrastructure with HTTP/WebDav support

The creation of an HTTP/WebDav ecosystem requires that all the storage elements support such kind of protocols and that the main parameters are tuned properly server side.

In the early stage testbed, we involved 14 Sites, distributed over the four different regions: Europe, North America, Asia and Australia. This wide distribution allowed us to test the protocol in a very realistic environment with different latency intervals. Additional aspects make the testbed highly relevant and representative, i.e. it cover the 60% of the Belle II production endpoints registered in the main Framework DIRAC, and includes systems configured with all the most popular storage technologies in High Energy Physics community i.e. : 5 dCache, 7 DPM and 2 STORM storage (table 1).
Table 1: List of sites included in the testbed

| STORAGE NAME | HOSTNAME | TYPE  |
|--------------|----------|-------|
| DESY-SE      | dcache-belle-webdav.desy.de | DCACHE |
| GRIDKA-SE    | f01-075-140-e.gridka.de      | DCACHE |
| NTU-SE       | bgrid3.phys.ntu.edu.tw       | DCACHE |
| SIGNET-SE    | dcache.ijs.si                | DCACHE |
| UVic-SE      | charon01.westgrid.ca         | DCACHE |
| Adelaide-SE  | coepp-dpm-01.ersa.edu.au    | DPM    |
| CESNET-SE    | dpm1.egee.cesnet.cz          | DPM    |
| CYFRONNET-SE | dpm.cyf-kr.edu.pl            | DPM    |
| Frascati-SE  | atlasse.lnf.infn.it          | DPM    |
| HEPHY-SE     | hephyse.oeaw.ac.at           | DPM    |
| Melbourne-SE | b2se.mel.coepp.org.au        | DPM    |
| Napoli-SE    | belle-dpm-01.na.infn.it      | DPM    |
| CNAF-SE      | ds-202-11-01.cr.cnaf.infn.it | STORM  |
| McGill-SE    | gridftp02.clumeq.mcgill.ca   | STORM  |

4. Belle II software with Davix Support

Basf2 is the Belle II framework for simulation and analysis, it is developed by the Belle II software group and it is based on the standard ROOT I/O libraries.
In order to support natively HTTP/WebDav URL in basf2, the TDavixFile.h library has been compiled in a testing machine, in substitution of TWebFile.h. After the first tests, the library has been included in the official release by the software development team since the build-2016-03-04.

The integration of TDavixFile library is a key step toward the creation of a fully HTTP/WebDav ecosystem, and to enable users to open HTTP URLs, transparently, in their steering files, with the standard basf2 syntax (figure 1).

```
filelistSIG=['https://belle-dpm-01.na.infn.it/dpm/na.infn.it/home/belle/TMP/test.root']
inputMdstList(filelistSIG)
```

Figure 1 – Example of file opening directive in a basf2 steering file.

Moreover TDavixFile provides specific parameters of ROOT that can affect the I/O performance, in particular from preliminary tests, we measured a large impact of the TTreeCache option. In figure 2, we show the time spent to read a file in streaming stored in the dCache Storage in DESY, and streamed from a user interface in Napoli, using an HTTP URL with and without TTreeCache.
Comparing the two graphs, we can see that for read the same 300MB file we spent 685 seconds with TTreeCache option switched off (left graph) vs 162 seconds with the feature enabled (right graph).
All the tests that will be presented in this work we have been performed with TTreeCache option enabled.
5. Test Plan

The main goal of our test plan is to estimate the impact of HTTP/WebDav protocols in the case of Belle II software, compared with the most used protocols in the HEP experiments: xrootd and gridftp. As reference benchmark, we have used a set of analysis jobs (basic skimming), in order to simulate as much as possible a realistic user activity. Test probes run within the Belle II framework basf2, and access to the input dataset with the ROOT I/O libraries. In order to highlight different aspects, two approaches have been considered, that reflect two possible analysis work-flows: Direct Data access and file download followed by a local read.

More specifically 5 different test probes have been prepared, each one doing a basic analysis on a single small file (10MB), in order to measure the protocol overhead rather than bandwidth effects.

The tests schema is the follow:

- Local input (download via HTTP)
- Local input (download via xrootd)
- Local input (download via gridftp+SRM interface)
- Input via HTTP streaming
- Input via xrootd streaming

All the testing probes use as input the same ROOT file, that has been preventively replicated over 12 of the 14 storage elements involved in the testbed, chosen among the more stable at benchmark time.
Testing jobs have been wrapped as Grid jobs, and then submitted to multiple Worker Nodes of the Belle II computing infrastructure. Each grid job performs the 5 test probes listed above against the selected 12 storage elements (figure 3). Each single probe is executed 10 times in order to have better statistics.

For each job we have collected the following information:

- Total time: Average, Max, Min, StdDev
- RootInput time: Average, Max, Min, StdDev
- Download Time: Average, Max, Min, StdDev
- Latency from the client

Network monitoring information have been retrieved obtained using the tool dstat.

6. Performance Analysis

In this paragraph we report the analysis that we did on the large performance's datasets, collected during the benchmark execution.

6.1. File download test

The first analysis that we did, aims at evaluate the download performance in function of the latency for each of the following protocols: HTTP, xrootd, and gridftp (through lcg-cp commands). The results reported in figure 4 show tests ran from a Worker Node in the RECAS-NAPOLI site [6][7], vs all the storage elements. In graph we have plotted the time in seconds spent to download a 10MB file by the analysis jobs that adopt the work-flow "file download + local read", for each SE and with different protocols. Storages are ordered by latency from the client.

We can state that HTTP, xrootd have identical behaviours in all the cases, while the test probes done with lcg-cp are systematically a bit slower, this is easily explained considering the overhead added by the SRM interface.

![Figure 4. File download performance from all the storage elements vs the same client running on a worker node on the RECAS-Napoli site.](image)

6.2. Remote Access

The second analysis report the performance for remote data access. More specifically we collected the time spent by ROOT I/O library, used internally by basf2, to stream the same 10MB file from different storages. The histogram in figure 5 represents the time spent to streaming the file from a worker node
installed in the McGill Site, from the DPM storages with HTTP and xrootd protocols. The storage are ordered by latency between client and SEs. Performance analysis in case of file streaming with dCache and STORM are still under investigation. The comparative analysis shown a substantial equivalence in protocol overhead in most cases, with a maximum discrepancy of 25% in a single case. The relative latency seems to affect the results much more than in the case of “download + local read” analysis strategy.

![Graph showing time spent by ROOT I/O library to stream the same file from all the DPM storages present in the testbed. The HTTP overhead is comparable with the xrootd in most cases.]

6.3. Remote Access

As complementary test, we performed the same bunch of probes with a 500MB file stored in the SE of RECAS-Napoli and read from a Worker Node in the same site. In this case we collected the time spent to complete the analysis job with different data access paradigm and protocols. The goal of this additional test is to verify HTTP performance with a lower latency incidence. The tests have been performed locally in order to create a safe and protected environment. As shown in the graph below (figure 6), the general behaviours emerged with the small file cases is maintained. More specifically: in case of analysis "download + local read" strategy, HTTP, xrootd and gridftp have basically the same performance. In case of analysis jobs with direct access (i.e. file streaming), HTTP shows an added overhead, this time under 1% to respect xrootd.

![Histogram representing the time spent to complete a single analysis job with different work-flow and with different protocols.]

Figure 5 - Time spent by ROOT I/O library to stream the same file from all the DPM storages present in the testbed. The HTTP overhead is comparable with the xrootd in most cases.
7. Conclusions

Belle II experiment is contributing at the international discussion on the usage of HTTP and WebDav in High Energy Physics experiments.
An extensive and representative testbed has been setup, which allows to test different aspects and e highlight the critical issues.
With the introduction of TWebDavix.h library in the basf2 software, we have proofed the possibility to work properly in full HTTP ecosystem with standard Belle II framework.
Performance analysis shown that in case of file download, HTTP has the same performance of a xrootd or SRM+gridftp storage, whatever the latency between client and server.
In case of data streaming, limited to the DPM storage, we report a small performance loss to respect xrootd that is highly optimized for ROOT and in this case the latency seems to play a not negligible role.
Additional tests will be done in the next future with dCache and STORM storages after the full optimization of the relative interfaces server side.

8. References

[1] Official webpage of the HTTP Deployement Task Force HTTPS://twiki.cern.ch/twiki/bin/view/LCG/HTTPDeployment
[2] Pardi S et al. “Testing and evaluating storage technology to build a distributed Tier1 for SuperB in Italy”- Journal of Physics: Conference Series 396 (2012) 042045 Proceeding CHEP12 Conference 21-25 May - New York (USA)- doi:10.1088/1742-6596/396/4/042045
[3] Furano F et al. “Direct data access protocols benchmarking on DPM “Journal of Physics: Conference Series 664 (2015) 042018
[4] Furano F et al. Towards an HTTP Ecosystem for HEP Data Access, J. Physics: Conf. Series 513, 032034 (2014). HTTP://stacks.iop.org/1742-6596/513/i=3/a=032034
[5] Pardi S et al. “Computing at Belle II” - Belle-II computing Group - Nucl.Phys.Proc. 273-275 (2016) 950-956 - DOI: 10.1016/j.nuclphysbps.2015.09.148
[6] Delprete D., Pardi S and Russo G “A Grid monitoring model over Network-Aware IaaS Cloud Infrastructure”- Journal: Int. J. of High Performance Computing and Networking, 2013 Vol.7, No.3, pp.195 – 204
[7] Palmieri F, Ficco M, Pardi S, Castiglione A- “A cloud-based architecture for emergency management and first responders localization in smart city environments” – Journal Computers and Electrical Engineering - Volume 56 Issue C, November 2016 Pages 810-830 doi>10.1016/j.compeleceng.2016.02.012

Acknowledgments
Authors wishing to acknowledge assistance of the staff of the ReCaS Data centre, the computing group of Belle II and the site admin of the Belle II production infrastructure.