WLCG space accounting in the SRM-less world

Julia Andreeva[1,*], *Dimitrios* Christidis[2,**], *Alessandro* Di Girolamo[1,***], and *Oliver* Keeble[1,****]

[1]CERN
[2]University of Patras

Abstract. The WLCG computing infrastructure provides distributed storage capacity hosted at the geographically dispersed computing sites. In order to effectively organize storage and processing of the LHC data, the LHC experiments require a reliable and complete overview of the storage capacity in terms of the occupied and free space, the storage shares allocated to different computing activities, and the possibility to detect "dark" data that occupies space while being unknown to the experiment's file catalogue. The task of the WLCG space accounting activity is to provide such an overview and to assist LHC experiments and WLCG operations to manage storage space and to understand future requirements.
Several space accounting solutions which have been developed by the LHC experiments are currently based on Storage Resource Manager (SRM). In the coming years SRM becomes an optional service for sites which do not provide tape storage. Moreover, already now some of the storage implementations do not provide an SRM interface. Therefore, the next generation of the space accounting systems should not be based on SRM. In order to enable possibility for exposing storage topology and space accounting information the Storage Resource Reporting proposal has been agreed between LHC experiments, sites and storage providers. This contribution describes the WLCG storage resource accounting system which is being developed based on Storage Resource Reporting proposal.

Introduction

The importance of the accounting of the computing and storage resources provided by the WLCG sites and used by the LHC experiments cannot be over-emphasized. Accounting is required for effective operations on the WLCG [1] infrastructure, for optimization of the resource usage and for strategic planning. CPU and storage accounting information represents important metrics for estimation of the quality of the service provided by the WLCG sites according to WLCG Memorandum of Understanding (MoU [2]).

Currently the CPU accounting for the WLCG resources is based on APEL [3] system for data collecting and processing and EGI Accounting Portal [4] for visualization. However,

[*] e-mail: Julia.Andreeva@cern.ch
[**] e-mail: dchristidis@ceid.upatras.gr
[***] e-mail: Alessandro.Di.Girolamo@cern.ch
[****] e-mail: Oliver.Keeble@cern.ch

the global WLCG storage space accounting is missing. Storage space accounting systems which have been developed by some of the LHC experiments, work in the scope of a single experiment and most of them are based on the SRM [5] protocol queries. Future WLCG evolution does not foresee mandatory deployment of the SRM service at all WLCG sites. Moreover, there is no plan to integrate new storage implementations with SRM. The new WLCG Storage Space Accounting System (WSSA) aims to provide accounting for disk and tape storage available on the WLCG infrastructure. The first implementation is based on the available data sources including SRM queries. Future evolution of the system foresees SRM-free approach which is described in Section 1.

In order to implement the WLCG Storage Space Accounting System, development should be driven in two directions:

- Enable exposure of necessary information by the primary information source, that is storage services
- Enable data collection, storage and visualization

Storage Resource Reporting (SRR) Proposal described in Section 1.1 addresses the first goal. Section 2 overviews implementation of the WLCG Storage Space Accounting Service.

1 Storage Resource Reporting (SRR)

1.1 SRR Proposal

Storage Resource Reporting Proposal describes the needs of the LHC experiments related to storage space information required for the computing operations. The requirements have been discussed with the experiments and the experts of the storage middleware development teams. Based on these requirements, a way to publish necessary information by the storage services has been proposed. This information consists of two parts, namely storage topology description and accounting data.

Storage topology description includes description of the storage shares and access protocols. Storage share is a distinct storage area dedicated for a particular use by a given LHC experiment. This term is equivalent to the space quota in SRM. Storage shares are accounted separately and do not have overlap in terms of space. Storage can be accessed by various protocols enabled on a given storage service. Protocols definition including endpoints has to be a part of the storage topology description in order to provide complete storage information required to exploit storage service. Storage topology description represents pretty static information. SRR foresees that storage topology description is provided in a JSON format and ideally should be accessible through the HTTP protocol. Though other protocols used by the LHC experiments for remote file access can be a valid option. The URL pointing to the topology description JSON file will be recorded in GOCDB [6]. It is foreseen that Computing Resource Information Catalogue [7] (CRIC) will periodically read storage topology description files and import data into CRIC.

In difference with the topology information, accounting data represents dynamic information. Accounting data defines the total capacity used and available to the experiment. SRR proposal suggests that accounting data should be updated with the frequency order of tens of minutes and accuracy of accounting data should be order of tens of GB. Such light requirements were agreed with the experiments. The goal is to provide required minimum and not to produce additional load on the storage services. There are two possibilities foreseen for getting accounting data:

- Query storage for free and used space with HTTP or XRootD protocols

- Publish accounting data in a JSON file which represents an extended version of the storage topology description complemented with accounting data

1.2 SRR implementation

SRR implementation by all storage flavours is progressing. For some storage flavours it is more advanced than for the others. For example, the DPM [8] releases starting from 1.10.0 enabled SRR. Therefore, the DPM sites are requested to upgrade to the latest DPM release and to enable configuration required for SRR. A dedicated task force has been set up to help sites with migration and re-configuration. Other storage middleware providers namely EOS [9], dCache [10], XRootD [11] and StoRM [12] have developed first prototypes for SRR implementation.

2 Implementation of the WLCG Storage Space Accounting Service (WSSA)

WSSA service collects storage topology and accounting information from the primary data sources, stores data in the repository and provides user interfaces. A set of APIs is being worked on.

2.1 Information sources

SRR implementation and deployment campaign might take long time, therefore, the initial implementation of the WSSA service exploits currently available data sources, which can depend on a particular experiment both for topology and accounting information. For example, LHCb and ATLAS are using SRM space quotas, therefore it was straight forward to use SRM queries for accounting information for these two experiments. For ATLAS topology, AGIS [13] provides complete topology description available through an API in JSON format. For LHCb topology, an XML file generated by DIRAC [14] is being used.

For ALICE, the WSSA service relies on MonALISA [15]. ALICE experiment developed storage space accounting system based on XRootD queries. However, XRootD queries not always provides correct info, therefore MonALISA applies some internal corrections. That is why it was decided not to use direct XRootD queries, but rather rely on information already collected and processed by MonALISA.

The only experiment for which disk storage accounting data is currently missing in WSSA is CMS. CMS does not use SRM disk quotas, it does not have an internal accounting system, therefore, CMS disk accounting data in WSSA will become available with SRR deployment at the WLCG sites.

2.2 Data storage

For data storage and visualization the WSSA relies on the MONIT [16] infrastructure which is used by many monitoring application at CERN. MONIT uses modern technologies like HDFS, Elasticsearch, InfluxDB and Grafana. It was straight forward to enable data import into MONIT and to create dashboard based on Grafana. Data is stored in Elasticsearch and InfluxDB, while Grafana dashboard uses InfluxDB as a storage backend.

2.3 Architecture

The WSSA service has been designed keeping in mind evolution of the information sources used by the system. As has been already mentioned above, in the first implementation, currently available data sources are used. As soon as complete WLCG storage topology description is provided by CRIC, it would be used as a source for topology data for all four experiments. With gradual deployment of SRR at the WLCG sites, SRR will replace SRM queries and MonALISA for the accounting data. No substantial changes in the WSSA system will be required to switch to a new data source.

Data flow for the initial implementation of the WSSA service is presented in Figure 1. Further evolution of the system is presented in Figure 2.

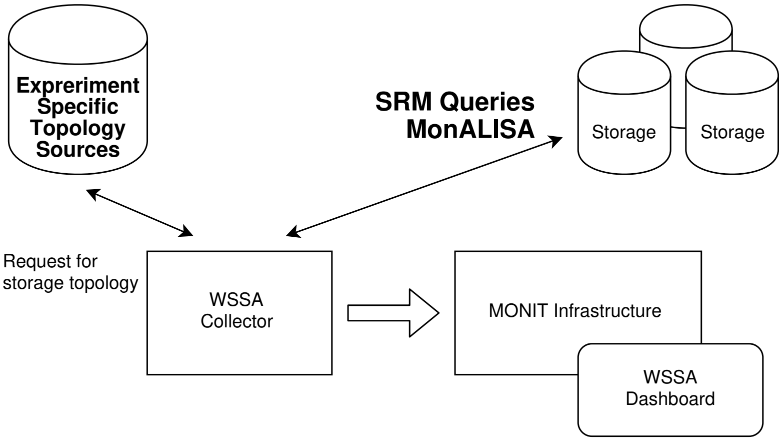


Figure 1. Initial implementation using the technologies available at the time.

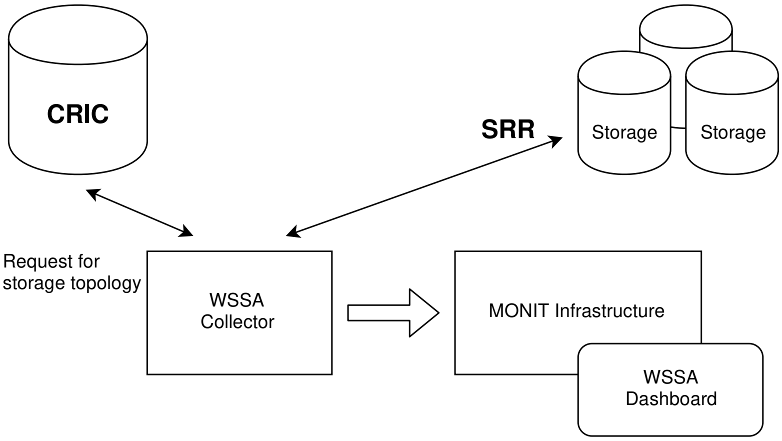
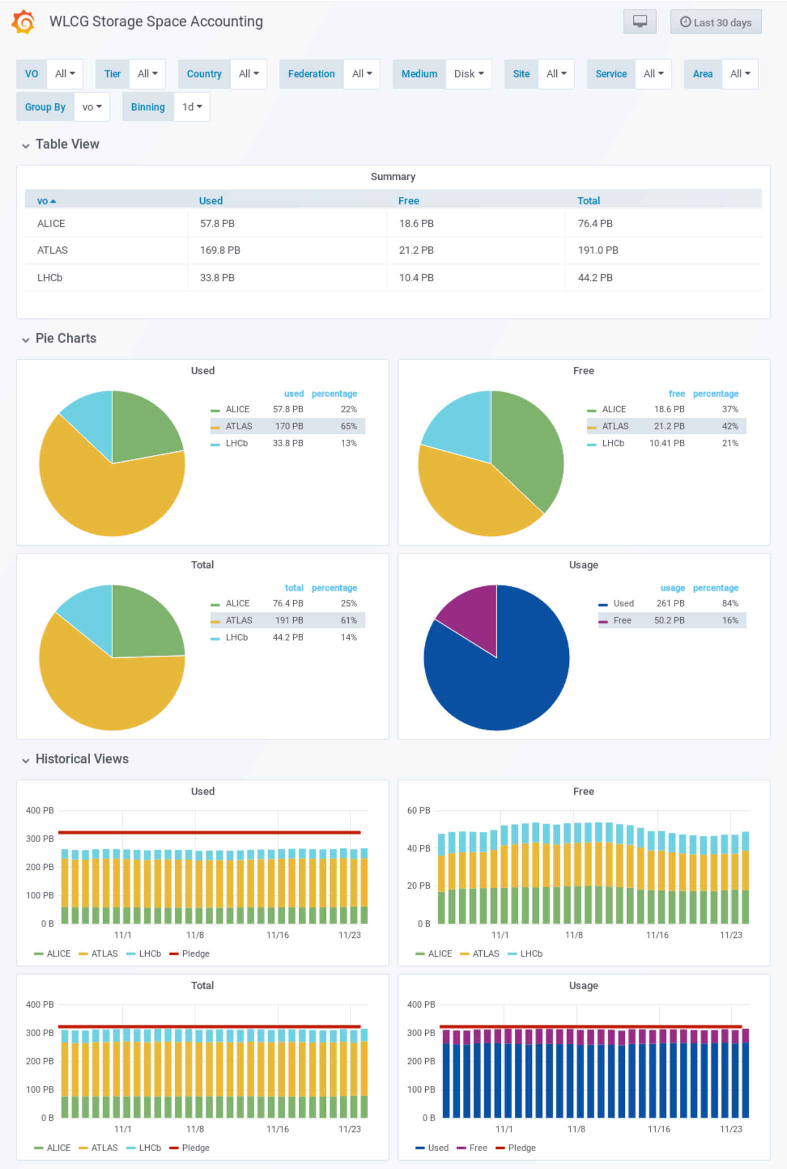


Figure 2. Expected data flow using CRIC and SRR.

2.4 User Interface

WSSA visualization is based on Grafana. WSSA User Interface shows distributions of storage accounting data over time as well as snapshots for the selected time range. Data on the UI can be grouped and filtered by experiment or LHC Virtual Organization (VO), tier, country, federation, type of storage (disk or tape), storage service, storage area or share. The example of the WSSA UI screenshot is shown in Figure 3.



WLCG Storage Space Accounting

Last 30 days

VO: All; Tier: All; Country: All; Federation: All; Medium: Disk; Site: All; Service: All; Area: All

Group By: vo; Binning: 1d

Table View

Summary

vo ▴	Used	Free	Total
ALICE	57.8 PB	18.6 PB	76.4 PB
ATLAS	169.8 PB	21.2 PB	191.0 PB
LHCb	33.8 PB	10.4 PB	44.2 PB

Pie Charts

Used

	used	percentage
ALICE	57.8 PB	22%
ATLAS	170 PB	65%
LHCb	33.8 PB	13%

Free

	free	percentage
ALICE	18.6 PB	37%
ATLAS	21.2 PB	42%
LHCb	10.41 PB	21%

Total

	total	percentage
ALICE	76.4 PB	25%
ATLAS	191 PB	61%
LHCb	44.2 PB	14%

Usage

	usage	percentage
Used	261 PB	84%
Free	50.2 PB	16%

Historical Views

Figure 3. The WSSA dashboard offers filters for fine-tuning and presents the information in multiple views.

3 Tape storage accounting

Tape storage accounting in the WSSA system has been enabled in collaboration with the WLCG Archival Storage Working Group. The goal of the group is to establish a knowledge-sharing community for those operating archival storage for WLCG and understand how to monitor usage of archival systems and optimise their exploitation by experiments. The first set of metrics published by the sites which run tape archives are being collected by the WSSA system and became available in the storage accounting dashboard as well as in the dedicated dashboard for tape metrics shown in Figure 4.

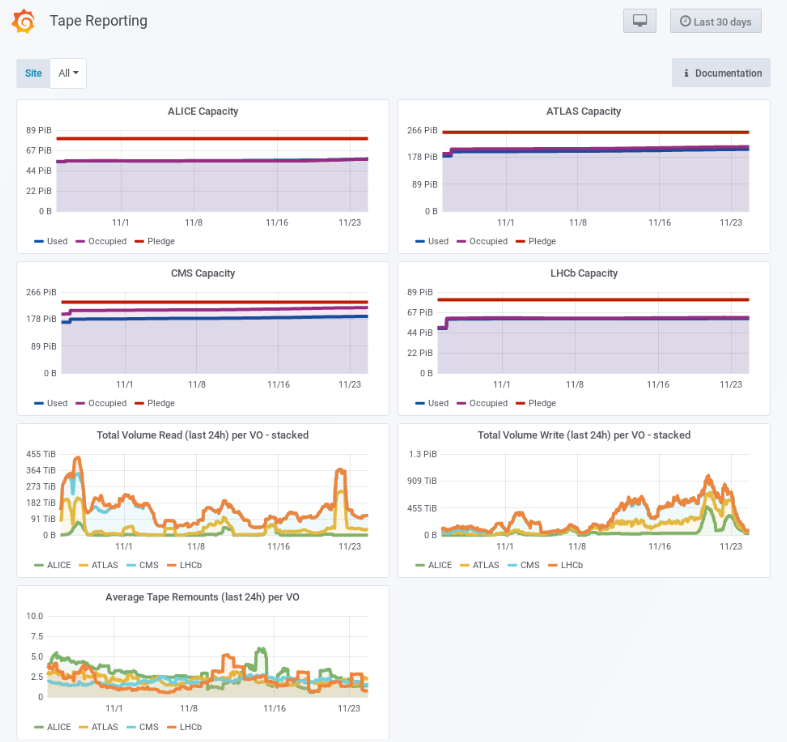


Figure 4. The tape-specific dashboard exposes additional metrics that are relevant to the tape experts.

4 Current status and plans

First version of the WSSA service is deployed in production. Currently it provides disk storage accounting for 3 LHC experiments: ALICE, ATLAS and LHCb, and for tape storage accounting for all LHC experiments. Next steps consist of gradual switch to the new information sources, as soon as they become available: CRIC for topology and SRR for accounting data.

Acknowledgement

This work represents result of collaboration of the WLCG operations, the MONIT team, storage middleware providers development teams, WLCG Archival Working Group and storage support teams at the WLCG sites. The authors would like to thank all people contributing to the development and deployment effort required to provide WSSA service.

References

[1] *Worldwide LHC Computing Grid*, accessed Nov. 2018, `http://wlcg-public.web.cern.ch/`

[2] *Memorandum of Understanding*, accessed Nov. 2018, `http://wlcg.web.cern.ch/collaboration/mou/`

[3] M. Jiang, C.D.C. Novales, G. Mathieu, J. Casson, W. Rogers, J. Gordon, *An APEL Tool Based CPU Usage Accounting Infrastructure for Large Scale Computing Grids*, in *Data Driven e-Science*, edited by S.C. Lin, E. Yen (Springer New York, New York, NY, 2011), pp. 175–186, ISBN 978-1-4419-8014-4

[4] *EGI Accounting Portal*, accessed Nov. 2018, `https://accounting.egi.eu/`

[5] *Storage Resource Manager*, accessed Nov. 2018, `https://en.wikipedia.org/wiki/Storage_Resource_Manager`

[6] *GOCDB*, accessed Nov. 2018, `https://goc.egi.eu/portal/`

[7] A. Anisenkov et al, *CRIC: a unified information system for WLCG and beyond*, EPJ Web Conf., (forthcoming)

[8] A. Manzi, F. Furano, O. Keeble, G. Bitzes, Journal of Physics: Conference Series **898**, 062011 (2017)

[9] A.J. Peters, L. Janyst, Journal of Physics: Conference Series **331**, 052015 (2011)

[10] *dCache*, accessed Nov. 2018, `https://www.dcache.org/`

[11] *XRootD*, accessed Nov. 2018, `http://www.xrootd.org/`

[12] *StoRM*, accessed Nov. 2018, `https://italiangrid.github.io/storm/`

[13] *ATLAS Grid Information System*, accessed Nov. 2018, `http://atlas-agis.cern.ch/agis/`

[14] *DIRAC*, accessed Nov. 2018, `http://lhcb-portal-dirac.cern.ch/DIRAC/`

[15] C. Grigoras, R. Voicu, N. Tapus, I. Legrand, F. Carminati, L. Betev, The European Physical Journal Plus **126**, 9 (2011)

[16] A. Aimar, A.A. Corman, P. Andrade, S. Belov, J.D. Fernandez, B.G. Bear, M. Georgiou, E. Karavakis, L. Magnoni, R.R. Ballesteros et al., Journal of Physics: Conference Series **898**, 092033 (2017)