

## **ASTRODS — DISTRIBUTED STORAGE FOR MIDDLE-SIZE ASTROPARTICLE PHYSICS FACILITIES**

**A. Kryukov<sup>1</sup>, I. Bychkov<sup>2</sup>, A. Mikhailov<sup>1,2</sup>, M.-D. Nguyen<sup>1</sup>, A. Shigarov<sup>1,2</sup>**

<sup>1</sup> *M.V.Lomonosov Moscow State University, Skobeltsyn Institute of Nuclear Physics.  
Leninskie gory, 1, bld.2, Moscow, 119992, Russia*

<sup>2</sup> *Matrosov Institute for System Dynamics and Control Theory, SB RAS.  
Lermontova, 134, box 292, Irkutsk, 664033, Russia*

E-mail: kryukov@theory.sinp.msu.ru

Currently, a number of experimental facilities for astrophysics of cosmic rays are being built and are already operating in the world. These installations produce large amounts of data that need to be collected, processed, and analyzed. Since many organizations around the world are involved in experimental collaboration, it is necessary to organize distributed data management and processing. Moreover, the widespread use of multi-messenger approach based on the coordinated observation, interpretation and analysis of disparate signals created by different astrophysical processes makes this problem much more urgent and complex. To meet a similar challenge in high energy physics, a WLCG grid was deployed as part of the LHC project. This solution, on the one hand, showed high efficiency, and, on the other hand, it turned out to be a rather heavy solution that requires high administrative costs, highly qualified staff and a very homogeneous environment on which applications operate. The paper considers a distributed data storage, called AstroDS, which was developed for middle-size astrophysical experiments. The storage currently integrates the data of the KASCADE and TAIGA experiments; in the future, the number of supported experiments will be increased. The main ideas and approaches used in the development of this storage are as follows: unification of access to local storages without changing their internal structure; data transfer only at the moment of actual access to them; search for the requested data using metadata and aggregation of the search results into a new collection available to the user. A distinguishing feature of the system is its focus on storing of both raw data and primary processed data, for example, data after calibration, using the Write-One-Read-Many method. Adding data to local repositories is carried out through special local services that provide, among other things, semi-automatic collection of meta-information in the process of downloading of the new data. At present, a prototype of the system was deployed on the basis of the SINP MSU.

**Keywords:** Distributed storage, Metadata, Aggregation service, Data life cycle, Astroparticle physics, Open science

Alexander Kryukov, Andrey Mikhailov, Minh-Duc Nguyen, Igor Bychkov, Alexey Shigarov

Copyright © 2019 for this paper by its authors.  
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

## 1. Introduction

Currently, a number of experimental facilities for astrophysics of cosmic rays are being built and are already operating in the world. Among them there are such installations as LSST [1, 2], MAGIC [3, 4], CTA [5, 6], VERITAS [7], HESS [8], and others. These installations produce large amounts of data that need to be collected, processed, and analyzed. Since many organizations around the world are involved in experimental collaboration, it is necessary to organize distributed data management and processing. Moreover, the widespread use of multi-messenger approach based on the coordinated observation, interpretation and analysis of disparate signals created by different astrophysical processes makes this problem much more urgent and complex.

To meet a similar challenge in high energy physics, a WLCG grid was deployed as part of the LHC project [9]. This solution, on the one hand, showed high efficiency, and, on the other hand, it turned out to be a rather heavy solution that requires high administrative costs, highly qualified staff and a very homogeneous environment on which applications operate.

Taking into account the tendency to analyze data from several sources simultaneously [10] for a more accurate investigation of the Universe and the modern movement towards Open Science [11], [12], it is very important to provide the users with access to data obtained from various astrophysical facilities. Since the amount of data obtained in some experiment often exceeds the capabilities of the relevant collaboration to process and analyze these data, the participation of all scientists interested in research in this area allows for a comprehensive analysis of the data in full.

Most existing collaborations have a long history and apply methods for data processing they are accustomed to. So, our approach to the design of data storage for astroparticle physics is based on two main principles. The first principle is that there is no interference with the existing local storage. And the second principle is the processing of user requests in a special service outside the local storage using the only metadata. Our approach to storage design is based on a single-write-multiple-read model (SWMR) for accessing data for further analysis. The motivation for the solution is that both raw data and data after the initial processing (for example, calibration) should be stored unchanged and presented to users as is. A similar approach is being discussed in the HDF5 community [13].

Thus the main ideas of the proposed approach are as follows:

- no changes in the inner structure of local storage;
- unified access to local storage based on corresponding adapter modules;
- use of local data access policies;
- search of the requested data using the only metadata on a special service;
- aggregating the requested data into a new collection and providing the user with access to it;
- data transfer only at the moment of actual access to them.

The prototype of such distributed storage, called AstroDS, is developed in the framework of the German-Russian Astroparticle Data Life Cycle Initiative [14]. The AstroDS currently integrates the data of the KASCADE [15] and TAIGA [16] experiments, which were chosen as a test case. Since the results of processing user requests for these test data have proved the viability, stability and efficiency of the developed storage, in the future the number of supported experiments will be increased.

The structure of the article is as follows. In Section 2, we discuss the basic components of AstroDS. Section 3 is devoted to the current state of work. In conclusion, we briefly discuss the main results and plans for the future.

## 2. THE STRUCTURE OF ASTRODS

The general structure of the AstroDS is presented on the fig.1.

The main ideas of the distributed storage is that we do not interfere with the work of local storages S1 ... S3. This is achieved by using special adapter programs A1 ... A3 that allow local storages to interact with the data aggregation service by unified manner. The data aggregation service is an intermediary program between the user and the rest of the system. As adapters, we use the CERNVM-FS [17] file system to export local file systems to the aggregation service in a read-only

mode. Using the CERNVM-FS file system results in data transfer only when the user actually requests these data. It should be noted that this important feature of CERNVM-FS significantly reduces network traffic.

The Metadata Catalogue (MDC) is a special service that determines which files contain data requested by the user. The MDC service is built around TimescaleDB [18,19].

To retrieve the necessary files, a user forms a request through the web interface provided by the Aggregation Service. When the Aggregation Service receives the user request, it first transmits the request in a modified form to the MDC, expecting to receive a list of all files containing the requested information and their location. After the Metadata Catalogue responds, the Aggregation Service forms the corresponding resulting response and delivers it to the user.

The AstroDS offers two types of search conditions for user requests: a file-level search and an event-level search.

In the case of a file-level search, the user forms a request containing conditions to be imposed on the data about the file as a whole. An example of such a condition is the range of dates of observation of cosmic rays. It is important to note that the user will receive in response the corresponding set of files with the same directory structure as in the original repository. Thus, the application software can be run without modification, as if the user runs the program locally.

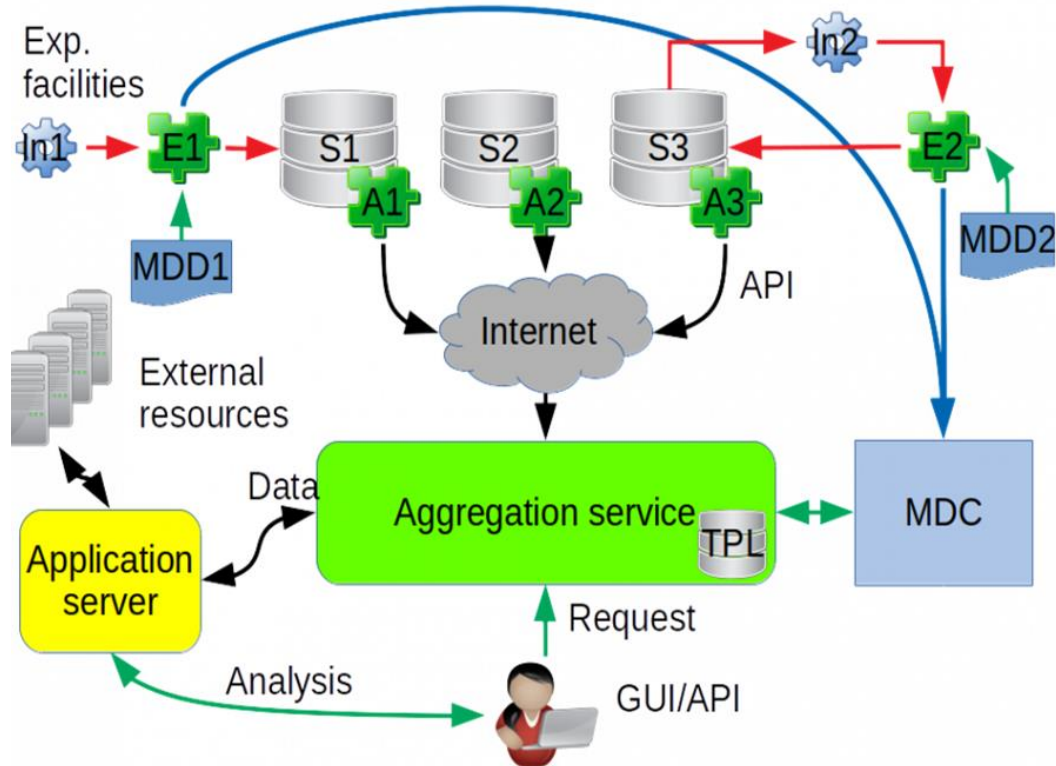


Figure 1. The structure of AstroDS

In the case of an event-level search, the user wants to select from the files only some events that satisfy the search conditions, for example, some energy range of the air showers. In this case the aggregation service selects only the necessary events and prepares a new collection consisting of these events, which is then transmitted to the user. The new set is transferred to the user. However, the directory structure will be preserved as well.

All data stored in the local storages must pass through the extractors. The extractors pick up metadata from the data and store the metadata in MDC. The type of the extracted metadata is defined by the metadata description file (MDD) which is used as input for the extractor. The MDD file is written in Kaitai Struct [20, 21] format with special marks pointing to elements of binary data which are metadata and should be extracted. In Fig. 1, extractor E1 extracts metadata from raw data, while extractor E2 extracts metadata during data processing, for example, from data after calibration or

shower energy). Thus, the information needed to process user requests is collected by the MDC service. For more details see [22]

Note that all services in AstroDS are built as microservices [23] and have a well-defined REST API. Some services work in Docker containers.

### **3. Current status**

Currently a prototype of AstroDS was deployed in Skobeltsyn Institute of Nuclear Physics, Lomonosov Moscow State University. The prototype consists of two local storages interconnected via a local network for modeling distributed storage, an aggregation service and a metadata catalog service based on TimescaleDB. The next version of the system will also include KCDC [24] storage at KIT and storage at Irkutsk State University.

Most of the components of the system are written in Python. As the first-time example of the production use of the system, users of the KASCADE and TAIGA/TUNKA collaborations will gain access to the data of these experiments, as well as the Monte Carlo simulation data. It should be mentioned that the system is developed for general use and is not limited to astrophysics applications.

### **4. Conclusion and future plans**

The first results of using AstroDS demonstrate the high functional characteristics of the system.

In the future, we plan to integrate more local storages with real data collections. To begin with, we plan to integrate the KASCADE, TUNKA, TUNKA-REX and TAIGA experiments. This will allow to begin trial operation of the system. A separate problem is the integration of data stored in the KCDC. The fact is that data in this system is not stored in the form of files, but in the database as separate records for each event.

Work is supported by RSF 18-41-06003. The authors also express deep gratitude to Yu.Dubenskaya for her help in preparing the article.

### **References**

- [1] Large Synoptic Survey Telescope // URL:<https://www.lsst.org/>
- [2] Kahn, S. M. Project Status // URL: [https://project.lsst.org/groups/sac/sites/lsst.org.groups.sac/files/Kahn\\_projectstatus.pdf](https://project.lsst.org/groups/sac/sites/lsst.org.groups.sac/files/Kahn_projectstatus.pdf)
- [3] MAGIC // URL: <https://doi.org/10.15161/oar.it/1446204371.89>
- [4] Ricoa, J. for the MAGIC Collaboration. Overview of MAGIC results // In. 37th International Conference on High Energy Physics, 2-9 July 2014 Valencia, Spain, Nuclear and Particle Physics Proceedings, 273275, 328-333 (2016)
- [5] Cherenkov Telescope Array. Exploring the Universe at the Highest Energies. // URL: <https://www.cta-observatory.org/>. Last accessed 24 Jan 2019
- [6] The Cherenkov Telescope Array Consortium: Science with the Cherenkov Telescope Array. // Arxiv: 1709.07997, URL: <https://arxiv.org/pdf/1709.07997>. Last accessed 24 Jan 2019
- [7] VERITAS // URL: <https://veritas.sao.arizona.edu/>. Last accessed 24 Jan 2019
- [8] HESS // URL: <https://www.mpi-hd.mpg.de/hfm/HESS/>. Last accessed 24 Jan 2019
- [9] Worldwide LHC Computing GRID // URL: <http://wlcg.web.cern.ch/>
- [10] Franckowiak, A.: Multimessenger Astronomy with Neutrinos. // J. Phys.: Conf. Ser., 888, 012009 (2017)
- [11] Voruganti, A., Deil1, Ch. , Donath, A., and King, J.: gamma-sky.net: Portal to the Gamma-Ray Sky. // Arxiv: 1709.04217, URL: <https://arxiv.org/pdf/1709.04217>. Last accessed 24 Jan 2019

- [12] Wagner, S.: Gamma-Ray Astronomy in the 2020s // URL: [https://www.eso.org/sci/meetings/2015/eso-2020/eso2015 Gamma Ray Wagner.pdf](https://www.eso.org/sci/meetings/2015/eso-2020/eso2015%20Gamma%20Ray%20Wagner.pdf). Last accessed Jan. 24 2019
- [13] HDF5 Single-writer/Multiple-reader. Users Guide // URL: [https://support.hdfgroup.org/HDF5/docNewFeatures/SWMR/HDF5 SWMR Users Guide.pdf](https://support.hdfgroup.org/HDF5/docNewFeatures/SWMR/HDF5%20SWMR%20Users%20Guide.pdf). Last accessed June 06, 2019.
- [14] Bychkov, I., et al.: RussianGerman Astroparticle Data Life Cycle Initiative // *Data*, 4(4), 56 (2018). DOI: 10.3390/data3040056.
- [15] W.D.Apel and etc. The KASCADE-Grande experiment // *Nuclear Instruments and Methods in Physics Research, Section A*, 620(2010), pp.202–216, DOI: 10.1016/j.nima.2010.03.147
- [16] Budnev, N. and etc. The TAIGA experiment: From cosmic-ray to gamma-ray astronomy in the Tunka valley // *Nuclear Instruments and Methods in Physics Research. Section A*, 845(2017), pp.330–333, DOI: 10.1016/j.nima.2016.06.041
- [17] Blomer, J. , Buncic, P. , Ganis, G. , Hardi, N. , Meusel, R., and Popescu, R.: New directions in the CernVM file system // In. 22nd International Conference on Computing in High Energy and Nuclear Physics (CHEP2016), 1014 October 2016, San Francisco, USA. *Journal of Physics: Conf. Series*, 898, 062031 (2017)
- [18] Freedman, M.J.: TimescaleDB: Re-engineering PostgreSQL as a time-series database. // URL: <https://www.percona.com/live/18/sites/default/files/slides/TimescaleDB-Percona-2018-main.pdf>. Last accessed 24 Jan 2019
- [19] Yang,Ch., et. al.: AstroServ: Distributed Database for Serving Large-Scale Full Life-Cycle Astronomical Data // ArXiv: 1811.10861. URL: <https://arxiv.org/pdf/1811.10861>. Last accessed 24 Jan 2019.
- [20] Kaitai Struct // URL: <http://doc.kaitai.io/>. Last accessed 24 Jan 2019
- [21] Bychkov, I. et al.: Using binary file format description languages for documenting, parsing and verifying raw data in TAIGA experiment. // In. International Conference "Distributed Computing and Grid-technologies in Science and Education" 2018 (GRID'2018), Dubna, Russia, September 10-14, 2018. *CEUR Workshop Proceedings*, 2267, 563-567 (2018).
- [22] Bychkov, I. and etc. Metadata extraction from raw astroparticle data of TAIGA experiment // In *Proc. of 3-d Int Workshop DLC-2019*.
- [23] Sill, A.: The Design and Architecture of Microservices // *IEEE Cloud Computing*, 3(5), 76-80 (2016)
- [24] KASCADE Cosmic Ray Data Centre (KCDC) // URL: <https://kcdc.ikp.kit.edu/>. Last accessed 24 Jan 2019