DATA KNOWLEDGE BASE: METADATA INTEGRATION SYSTEM FOR HENP EXPERIMENTS

M. Golosova[1], M. Grigorieva[2], V. Aulov[1], A. Kaida[3], M. Borodin[4]

[1] *NRC "Kurchatov Institute", 1 Akademika Kurchatova sq., Moscow, 123182, Russia*

[2] *Lomonosov Moscow State University, 1 Leninskie Gory, Moscow, 119991, Russia*

[3] *NR Tomsk Polytechnic University, 30 Lenina prospekt, Tomsk, 634050, Russia*

[4] *University of Iowa, Iowa, USA*

E-mail: golosova_mv@nrcki.ru

HENP experiments, especially the long-living ones like the ATLAS experiment at the LHC, have a diverse and evolving ecosystem of information systems that help scientists to organize research processes – such as data handling (including data taking, simulation, processing, storage, and access), preparing and discussion of publications, etc. With time all the components of the ecosystem grow, develop into complex structures, accumulate metadata and become more independent and less flexible. Automated information integration becomes a pressing need for effective operation within the ecosystem. This contribution is dedicated to the meta-system, known as Data Knowledge Base (DKB), designed to integrate information from multiple independent sources and provide fast and flexible access to the integrated knowledge. Over the last two years, the system is being successfully integrated with the production system of the ATLAS experiment, including the extension of the production system web-interface with functionality built upon the unified metadata provided by DKB.

Keywords: information integration, online analytics, metadata, mega-science

Marina Golosova, Maria Grigorieva, Anastasiia Kaida, Vasilii Aulov, Mikhail Borodin

1. Introduction

Present-day HENP scientific experiments at the forefront of research often involve complicated facilities (like LHC [1] or NICA [2]), massive apparatus (like ATLAS [3] detector at LHC or XFEL [4]) and are going on for years and even decades. Every project of such scale (mega-science project) must be accompanied by some information infrastructure to manage all operations of a research process. Being unique for every particular project, each infrastructure addresses similar set of tasks: experimental data storage, search and access, modelled data production, software development, execution of computing tasks on shared computing resources, resources allocation for different scientific groups, etc. For every type of activity, executed operations often produce and/or require some auxiliary information about data or processes – metadata. These metadata are also used for monitoring purposes and to get current state of specific processes. Combined together, different types of metadata can additionally be used to get summary information about the whole infrastructure – in order to detect possible or actual problems, find ways to solve them and determine directions for the further development.

The more complicated the project is and the more diversified tasks the infrastructure must serve, the more complex is the metadata management. In very simple cases metadata may be managed by a single information system, taking care of consistent storage, operative updates and providing users and analytics with convenient representation of the information for every use-case; yet for mega-science project the volumes and diversity of metadata lead to development of significantly independent systems, operating with different scopes of metadata. These metadata scopes, even being related to various parts of the project, still remain semantically connected – and for analytical purposes often appear to be useful to bring them together and treat information from multiple scopes in terms of some unified metadata model. In case of a diverse infrastructure, where wide variety and vast volumes of information are handled by multiple different systems, this task requires special attention, as straightforward information integration and aggregation "on the fly" may take a lot of time, making interaction with the analytical tools strictly offline. Offline analytics is widely used for regular pre-defined reports generation, but it is inefficient for analytical research on the infrastructure operation as a whole.

This paper provides the authors' view on the problem of the interactive analytics for complex information infrastructures of HENP experiments and describes an approach applied to the development of the meta-information system prototype, aimed to serve multi-scope analytical tasks, in the case of the ATLAS experiment at LHC.

2. Metadata concept hierarchy

In the case of the ATLAS experiment at LHC, the online analytics is mostly interested in the metadata related to the physics data production (modelling and preparation for analysis), storage and analytical processing. The high-level management is performed in global terms, such as "Monte Carlo simulation campaign" or "data sample", but metadata are managed on a level of smaller objects – such as "dataset" (data storage unit, defined in Rucio meta-catalog [5]), "production request" (generated by user or group request for massive data processing) or "task" (processing unit, defined in Production System 2 [6]). The high-level objects are mostly defined in human-readable form and used by people to simplify conversation; they carry valuable semantic charge but very little metrics information useful for analytical processing. On the contrary, the low-level objects are less semantically charged, yet provide plenty of metrics for events and objects, directly managed by corresponding systems.

Figure 1 represents the concept hierarchies for both data storage and computing in terms of the ATLAS information infrastructure; it also provides examples of metrics, related to specific concepts and links to the systems that manage this information: JIRA, DEFT, JEDI, AMI and Rucio.

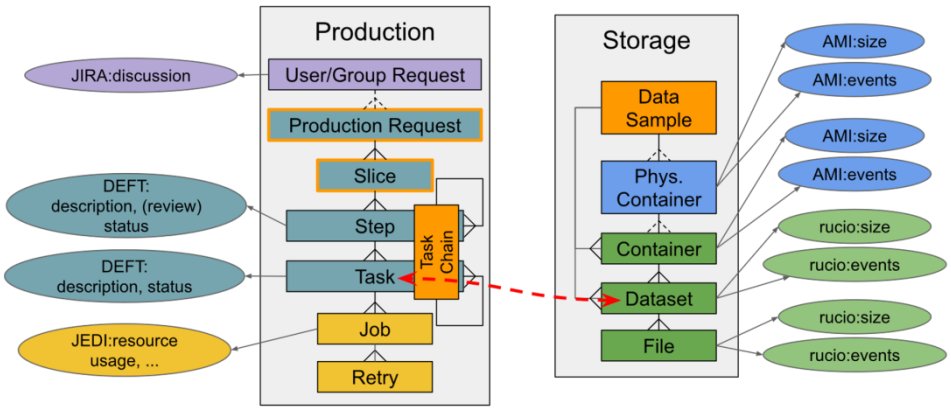

Figure 1. Metadata concept hierarchies and data sources (in the case of the ATLAS experiment)

The most common analytical request is to calculate some low-level metric for a high-level object. In terms of the ATLAS information infrastructure, the requests can be, for example, to calculate the storage size of a *Data Sample* (as a sum of storage sizes of all included *Datasets*), or CPU resources spent on its production (as a sum of the CPU usage metric of all *Jobs*, executed to produce included *Datasets*, at every processing *Step* of *Task Chains* starting from the first *Step* (modelled data generation or event reconstruction for the real data taken from the detector) – or from any other specific one).

3. Data Knowledge Base

The Data Knowledge Base meta-information system is a system designed to bring together different scopes of metadata, reconstruct missed or indistinct links between objects and provide fast and flexible access to information, in particular – about the high-level objects. The prototype, developed for the Production System [7] of the ATLAS collaboration, uses Elasticsearch full text search engine [8] to store integrated metadata at the level of *Task* and *Dataset* objects. The data model used to index information about these objects presumes that *Task* object properties include the properties of the *Task* itself, references to the higher level objects (like *Task Chain* or *Campaign*), and some properties aggregated by the low-level objects (*Jobs*) – such as actual CPU usage, for example – while the *Dataset* properties contain only the *Dataset* object properties (both storage information and data characteristics) and no additional information. Each object within the Elasticsearch is represented as a document (with object properties as the document fields), and *Task*/*Dataset* documents are connected as parent/child entities, where *Task* is a parent, and its output *Datasets* are child documents.

This indexing model described above made it possible to implement metadata integration as a two-branched ETL (Extract, Transform, Load) process, where one branch is responsible for integration and indexing metadata of a *Task* object, and the other one processes *Dataset* objects related to the *Task*. However, this division of information into two different types (*Task* and *Dataset* metadata), while making the prototype development process simpler, has also imposed some restrictions on the metadata usage scenarios. And the implementation of real-life use-cases for the Production System users (production managers) revealed that these restrictions do affect the response time of the prototype for user requests, for in many scenarios *Dataset* properties are treated as those of the *Task* object. These scenarios require additional efforts for information retrieving, making the request execution less performant. Depending on the number of the documents in the selection, some of the implemented requests might take tens of seconds – while for the seamless interactive communication with the system, the response time should not exceed 10 seconds [9].

4. Indexing model improvement

To improve the performance of user operations, a new metadata indexing model was suggested. It takes into account the specifics of the already addressed use-cases, and the most noticeable change is that the output datasets properties are now stored together with the Task object, in the form of nested documents (instead of parent/child documents). Operations with nested documents show better performance due all related documents being stored not only in the same shard, but in the same Lucene block. It also means that re-indexing of one of the documents leads to re-indexing of all related ones – but when, as in this case, indexed documents contain only object metadata (and not the object content, which may be sizeable), their sizes are quite small and it will not be very expensive operation.

To test the new model against the one originally used in the prototype, a specially allocated single-node instance of the Elasticsearch was used (heap space: 4GB; index volume: 4M records (~2M tasks, ~2M datasets) (5GB)). During the testing all caching mechanisms were disabled, and both ES and disk caches that could not be disabled were cleaned after every request: reading data from memory – and even more, getting request results from the cache – would almost eliminate the effect of scheme change on given data volumes, reducing the request execution time to almost immediate response.

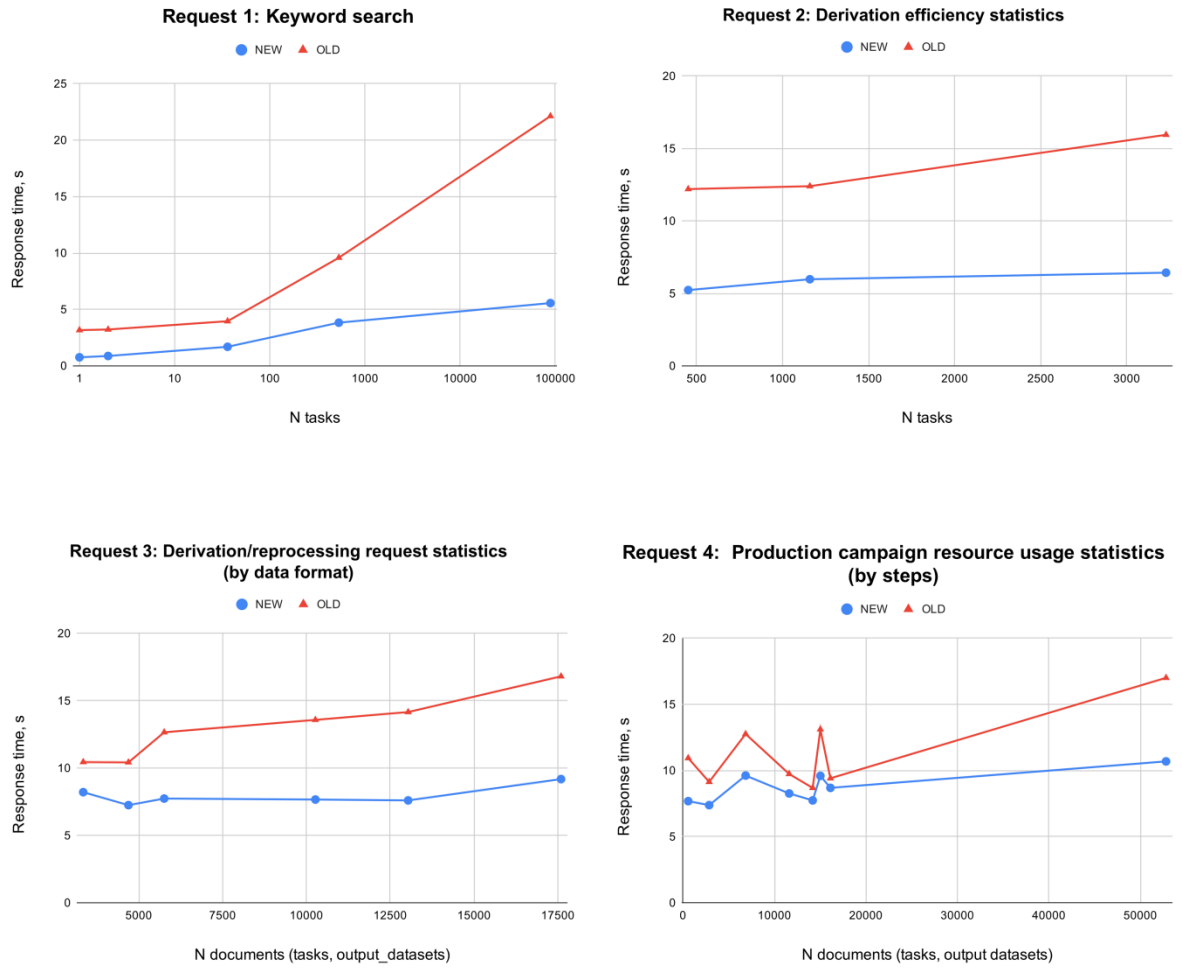

Figure 2. Test results: response time in relation to number of tasks or documents (both tasks and output datasets) matching the requests 1-4: (1 - keywords search, 2 - calculation of derivation process efficiency, 3 - calculation of general statistics for derivation or reprocessing request by output data format, 4 - calculation of general statistics for campaign by processing steps)

The comparison of two indexing models was made in 4 different cases:

- keywords (Google-like) search of tasks and related datasets;
- calculation of derivation process efficiency statistics in accordance with the output data format (aggregation by all input and output datasets of tasks selected by specific properties);
- general statistics (resource usage, number of processed events, …) for:
    - MC production campaign (by steps);
    - reprocessing or derivation request (by output data format).

The test results (Fig. 2) show that new indexing model reduces the response time in all test cases and also is more scalable. It means that introducing this model into the prototype will allow users to perform interactive analysis of greater volumes of metadata with the same configuration of the backend Elasticsearch cluster, reducing the resource requirements.

5. Conclusion

The Data Knowledge Base prototype, successfully integrated with the Production System of the ATLAS experiment, allows execution of complex analytical requests, requiring information from different information systems and from different levels of abstraction. Some metadata in use may belong to a computing *Job*, executed on a single computing node, while another describe a *Data Sample*, which may include petabytes of data produced under the same conditions (software version, configuration parameters, etc).

Although the prototype allows implementation of multiple different scenarios, providing flexible access to the integrated metadata, in some cases the response time still exceeds the upper limit usually considered for the interactive operations. Possible solution is to improve the metadata indexing and storage scheme, bringing pieces of information that are often used together into a single document (with nested sub-documents).

Performance testing of both schemes (currently used one and the one developed according to a suggested solution) has proved that the suggested changes in storage scheme do improve the performance of user operations and also make the system more scalable. Currently the DKB development team works on the update of the operating DKB prototype installed at CERN to apply these changes: in addition to re-indexing of all the stored data with the new scheme, all the metadata integration scenarios (ETL processes), responsible for the filling and regular update of the Elasticsearch storage, are being updated accordingly; the update also requires improvement of user interface to make this (and any other possible in the future) change in the storage scheme transparent for the end users.

6. Acknowledgement

This work is supported by Russian Science Foundation under contract №18-37-20003.

References

[1] LHC, Large Hadron Collider // CERN Publication, European Laboratory for Particle Physics, June 1990

[2] Agapov N. et al. Design and Construction of Nuclotron-based Ion Collider fAcility (NICA). Conceptual Design Report, edited by I.Meshkov and A.Sidorin // JINR, Dubna, 2008. Available at: http://nica.jinr.ru/files/NICA_CDR.pdf (accessed 06.11.2019)

[3] ATLAS Collaboration. The ATLAS Experiment at the CERN Large Hadron Collider // JINST 3 S08003, 2008. Available at: http://nordberg.web.cern.ch/nordberg/PAPERS/JINST08.pdf (accessed 06.11.2019)

[4] Altarelli M. et al. XFEL: The European X-Ray Free-Electron Laser – Technical Design Report // DESY 2006-097, 2006. Available at: https://xfel.desy.de/localfsExplorer_read?currentPath=/afs/desy.de/group/xfel/wof/EPT/TDR/XFEL-TDR-final.pdf (accessed 06.11.2019)

[5] Rucio – Scientific Data Management // https://rucio.cern.ch (accessed 06.11.2019)

[6] Barreiro F.H., Borodin M., De K., Golubkov D., Klimentov A., Maeno T., Mashinistov R., Padolski S., Wenaus T. on behalf of the ATLAS Collaboration. The ATLAS Production System Evolution: New Data Processing and Analysis Paradigm for the LHC Run2 and High-Luminosity // IOP Conf. Series: Journal of Physics: Conf. Series 898 (2017) 052016, doi :10.1088/1742-6596/898/5/052016. Available at: http://inspirehep.net/record/1638474/files/pdf.pdf (accessed 10.10.2019)

[7] Golosova M.V., Aulov V.A., Grigoryeva M.A., Kaida A.Y. Data Knowledge Base for the ATLAS collaboration // Proceedings of the VIII International Conference "Distributed Computing and Grid-technologies in Science and Education" (GRID 2018), Dubna, Moscow region, Russia, September 10 - 14, 2018

[8] Elasticsearch // https://www.elastic.co/products/elasticsearch (accessed on: 17.10.2018)

[9] Nielsen J. Usability Engineering // New York: Academic Press, 1993