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

DATA KNOWLEDGE BASE FOR THE ATLAS COLLABORATION

M.V. Golosova [1,a], V.A. Aulov [1], M.A. Grigorieva [1], A.Y. Kaida [2], on behalf of the ATLAS Collaboration

[1] *Laboratory of Big Data Technologies for Mega Science Projects, NRC «Kurchatov Institute», 1 pl. Akademika Kurchatova, Moscow, 123182, Russia*

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

E-mail: [a] golosova_mv@nrcki.ru

ATLAS experiment at the CERN LHC is one of the most data-intensive modern scientific apparatus. To manage all the experimental and modelling data, multiple information systems were created during the experiment's lifetime (more than 25 years). Each such system addresses one or several tasks of data and workload management, as well as information lookup, using specific sets of metadata (data about data). Growing data volumes and the computing infrastructure complexity require from researchers more and more complicated integration of different bits of metadata from different systems using different conditions. A common problem is multi-system join queries, which are not easy to implement in a timely manner and, obviously, are less efficient than a query to a single system with integrated and pre-processed information would be. To address this issue, a joint team of researchers and developers from Kurchatov Institute and Tomsk Polytechnic University has initiated the Data Knowledge Base (DKB) R&D project in 2016. This project is aimed at knowledge acquisition and metadata integration, providing fast response for a variety of complicated queries, such as finding articles based on same or similar data samples (search by links between objects), summary reports and monitoring tasks (aggregation queries), etc. In this paper we will discuss main features and applications of the DKB prototype implemented by now, its integration with the ATLAS Workflow Management, and future perspectives of the project.

Keywords: metadata integration, information integration, metadata, ETL, multi-source query

1. Introduction

In modern times major scientific experiments take form of international long-living projects with unique work flow and infrastructure of information systems. ATLAS experiment at the CERN LHC [1], one of the most data-intensive modern scientific apparatus, is one of such projects. Growing data volumes and the computing infrastructure complexity lead to the necessity of multiple whiteboards, dashboards and monitoring pages that provide detailed and/or summarised representation of the project subprocesses and subsystems (e.g. trigger whiteboard, data quality whiteboard, MC production campaign overview, DDM storage usage plots and tables) [2]. Usually, these whiteboards (and the processes they are representing) are loosely connected to each other and in every given use case their development goes independently, sometimes solving similar problems (e.g. of scheduling and failure recovery mechanism) multiple times.

The Data Knowledge Base R&D project, aimed at enhancement of the connectivity between different areas of the community activities, was initiated by a joint team of researchers and developers from Kurchatov Institute and Tomsk Polytechnic University in 2016.

During the first year of the project, briefly discussed in the first section of this paper, the team faced challenges of information integration from different sources, similar to those being solved by multiple already existing whiteboards. Due to the research nature of the project and the fact that the DKB was supposed to work with wide variety of information sources, a flexible way to solve mentioned problems of multi-source queries became one its goals. Suggested approach proved to be valid for the developed prototype, leading to the DKB conceptual extension (DKB v.2). The information integration problem, achieved results and addressed by DKB v.2 real-life use cases are discussed in the second section. It also provides authors` view on the possible directions of the further development of the project. The experience gained in this work is summarised in the conclusion.

2. Data knowledge base

One of the questions arising from the loose connectivity between different subprocesses of the whole experiment work flow is as follows: which of the data, produced by the ATLAS detector or generated with Monte Carlo simulation, were used in studies that led to publications? In other words, which data should never be removed to provide reproducibility of the published results?

Within the ATLAS collaboration multiple research activities are being conducted simultaneously. Each study is based on physics data (information about particle collision events), stored in data samples in Grid centres all over the world. With time new samples are added, while some of the stored ones can become obsolete (e.g. due to the new software version release). However, due to the essential requirement of reproducibility of scientific results [3], data used in physics analysis should be kept as long as possible — ideally, forever. Yet storing all the data forever would take too many resources (volume of the stored data in ATLAS is more than 380 petabytes and keeps growing); so selection of data used for analysis become vital to ensure preservation while having a possibility to remove obsolete and unused samples. Since 2016, this information can be explicitly provided by authors via the interface between ATLAS Metadata Interface (AMI) [4] and Glance systems [5]. But for the publications made before this functionality became available, and for some part of the works published after it (as long as providing this information is not mandatory), these data are mostly available in human-oriented resources: internal documents, web-pages created by researchers, spreadsheets, data requests and discussions in dedicated systems (e.g. Indico, Jira, Twiki).

The project of DKB was initiated as an attempt to extract this information in the automated way, without human intervention, reconstructing connections between publications and all the information available about the research process. Conceptual architecture and prototype development are beyond the scope of this paper and discussed at full length in [6] and [7].

The first year results of the DKB project allowed to say that valuable metadata can be extracted from unstructured sources and used, for example, to reconstruct connections between public documents and collaboration internal processes. These connections can be combined into a web, that ties together objects not associated directly (basing only on the metadata available for the objects themselves). The developed prototype can be used as a basis for a fully functional system, that allows users to navigate this web to find, for example, information related to the same topic (publications, documents, discussions), or studies based on the same data. [7]

At the same time, the metadata integration system developed as a part of the prototype to integrate information from arbitrary resources (both structured and unstructured) looked very promising as a starting point for many other tasks requiring data from multiple independent sources. This observation gave birth to the DKB v.2, discussed in the next section.

3. Metadata integration

3.1. Hypothesis

The mentioned information integration and aggregation task is very common in such a long-living project as the ATLAS experiment. There are two general ways to solve it in every specific use case: a direct query (sometimes multi-source — a number of queries to different systems with join operation on the client side) and periodic creation of the query results snapshot. First one is very flexible and works good enough unless there are too many data to be integrated, but when the direct queries becomes too heavy (e.g. production and analysis jobs monitoring and inspecting, with number of jobs reaching 2M per day) [8], pre-created snapshots can be a better solution. They allow providing more or less actual information very fast, however flexibility is limited: a snapshot can hardly be reused for different use cases, even based on the same set of raw metadata.

In the real world there are situations requiring the combination of these two approaches: pre-created static snapshots are used to get the general information and assess if more detailed information is needed, while the detailed information is taken directly from the respective systems. An illustration of such a scenario can be production coordinators` actions performed to detect and solve possible issues. High-level summary tables (Twiki pages or ProdTask [9] interface) provide current state of the production campaign; if the summaries reveal problems, the next step is to find failed computing tasks via BigPanDA [10] or ProdTask. Then, depending on the failure type, information from multiple other sources may be needed to find out the reason and ways to recover. Most of the actions require manual interaction with Web GUIs, specific command line clients, database clients, etc., and responses from all the systems involved may take quite a lot of time. A possibility to avoid querying all the systems could significantly simplify the scenario described above; to achieve it, all the possibly needed information should be joined by key fields and properly organised in some storage to provide reasonably fast response for the whole variety of queries used in every given use case.

Metadata integration system, suggested for the DKB v.1, provides a basis for this task; to use it, for every query family (a group of queries sharing information field area and query type) the following steps should be performed:

- choose the most appropriate data storage and access technology (relational database, full-text index, object storage, etc.);
- create the data scheme that maps the original metadata structure to the chosen storage and fits metadata usage requirements;
- organise and schedule the ETL process (data extraction, transformation and load) from the original sources to the chosen dedicated storage according to the addressed use cases requirements;
- implement query or queries providing the required information.

The main idea is not to create a single storage fitting all the use cases, which is difficult to achieve and even more difficult to keep up to date (as any new use case may require changes), but to build a system that would allow addressing new use cases with minimal effort by adding modules to it.

3.2. Integration with the ATLAS Production System

The concept described above was applied to the use cases of the ATLAS Production System monitoring and control interface ProdTask. All the use cases taken for the DKB v.2 prototype development fell in two main categories:

full text search of production and analysis tasks based on all the text metadata available for the tasks and related data samples (input and output data);

aggregation of a number of parameters (e.g. number of events in the input and output data samples) by other parameters (e.g. campaign to which the task belongs) with given requirements.

These categories share the information field area (computing tasks and related data samples), yet have different query types (full text search and aggregation). As a back-end technology that would allow addressing both of them, the Elasticsearch [11] full text index and search system was chosen. It provides flexible full text search with the possibility to assign custom analysers to specific fields (useful for fields obeying specific nomenclature, such as data sample names [12]) along with powerful aggregation and analytic capabilities.

However, as any Big Data oriented technology, Elasticsearch has several limitations. Being a very powerful instrument for operations with documents of the same type (where "document" is an ES specific representation of an object, and "document type" is a logical category; documents in the same category are usually expected to have similar structure), it is very limited in terms of relations between objects (especially for many to many relationship). In the use cases under consideration data model was quite simple: primary input data sample → (1:1) computing task → (1:N) output data sample. To minimise restrictions on possible queries, all the objects related to one task could be joined into a single document (e.g. via nested documents), but it would make integration process more complex: as most of Big Data oriented storages, Elasticsearch does not support update operation (as it is very expensive in terms of performance), meaning that the whole document should be written at once. In order to have possibility to run ETL process related to a single output data sample metadata independently from that related to the whole set of task metadata (including all output data samples metadata), it was decided to make the separation on the data scheme layer: one document type for information about task and input data sample, another – for output sample [13]. This scheme also reflected the structure of original metadata sources in some way — not an essential requirement, but allowed to separate integration process in two semi independent processes that share only the first processing stage — initial data extraction form the Production System (see Figure 1). In other words, the development process was simplified at the cost of possible difficulties with new queries implementation in order to get the working prototype as soon as possible. None of the use cases in the consideration suffer from this decision, but if it happens in the future, the scheme should be changed. For the real world, not prototype, applications it is better to prefer the data scheme that can serve the widest diversity of query types early in, even if it means to have a more overloaded integration process and spend more time on its development.

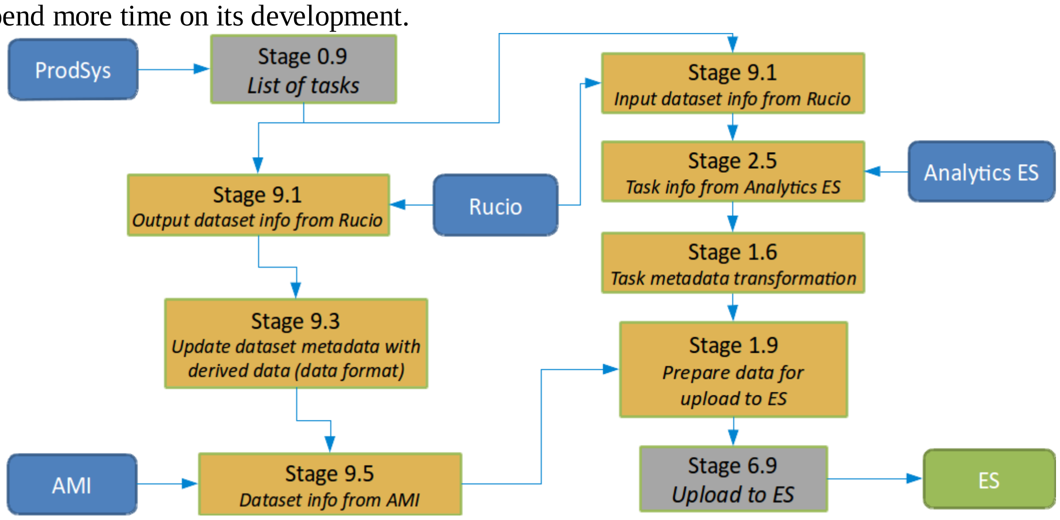

Figure 1. ETL process for ATLAS data production and analysis tasks metadata integration

- The list of metadata sources in use is extendable: when additional information is needed, new sources could be added, just as Rucio [14], AMI and CloudLab Elasticsearch were added at some point.
- Currently the DKB Elasticsearch index contains metadata of all the tasks and related datasets since 2014 (about 3.6M documents) and is regularly updated. Hourly update takes ~2 min, meaning that it can be scheduled to run more often. The full reprocessing (sometimes required even in the real world applications, e.g. to add new data source) takes quite a lot of time for now: ~20 hours for a single month metadata. To make new version of the integrated data available as soon as possible, it was organised to start from the most recent (and most actual) metadata and go to the oldest archive record. The situation can be improved by using processing parallelisation capabilities of Apache Kafka [15] being in the basis of the integration system; it was not done yet, but is planned for the future.
- The integrated metadata are used for a number of web pages with the following functionality:
- tasks full text lookup;
- statistics for derivation tasks (related to given parametric values of a project and AMI tag);
- campaign statistics: aggregated by steps information about output and input/requested data.

3.3. Future plans

Along with the further development and improvements of the DKB v.2 core, responsible for the metadata integration processes, there are also plans of usability improvements for the already existing web interfaces (e.g. to extend the full-text task search with search by selected fields), and addressing of new use cases requiring different functionality (e.g. task chain reconstruction, that requires search by links between objects; graph databases look like a possible solution and will be evaluated).

Taking into account the fact that internal storages of the DKB should evolve with the use cases being addressed, an abstraction level of API will be added to hide the internal structure from the clients. It will allow making changes in the internal structure in order to improve performance of the already addressed queries and implement new ones without the need to update the client side every time when the data scheme or even back-end technology is changed.

4. Conclusion

- The R&D project of the Data Knowledge Base, initiated in 2016 as an attempt to enhance connectivity between different areas of the ATLAS community activities, has evolved into a system that simplifies access to the metadata originated in different information systems in both structured and unstructured form. It provides tools for information integration from multiple sources of different types into dedicated storages, allowing fast and flexible access to the information field of the ATLAS experiment as a whole.
- First prototype allowed to integrate information about projects and campaigns from AMI, about published papers and supporting notes (internal documents) from CDS [16] and Glance, and to extract from supporting documents themselves additional information that links mentioned entities together.
- Approach to metadata integration utilised in that prototype is case independent and can be used for different scenarios; it is aimed at the simplification of the common tasks of metadata integration — organisation and management of ETL processes. It was used for real life use cases of the ATLAS Production System and has already allowed to extend the functionality of ProdTask interface with the full text search and aggregation capabilities of Elasticsearch storage, applied to information collected from Production System (DEFT, JEDI) [17], AMI, Rucio and ATLAS Analytics CloudLab Elasticsearch cluster. This functionality is meant to simplify daily operations of production coordinators, reducing the number of actions to be performed to get the necessary information for production process issues detection and solving.
- The prototype can be extended in different ways:
- adding new metadata sources to the already established integration process;

- adding new integration processes, based on a completely different set of metadata sources;
- adding new internal storages, optimised for new types of queries.
- Each of them is planned to be used accordingly:
- if currently addressed use cases of ProdTask will require new data sources (as it was already done for CloudLab Elasticsearch when information about task CPU time was required);
- when new use cases, not related to the production and analysis tasks, appear;
- to evaluate possibility and advantages of task chain reconstruction via graph database.

Acknowledgment

This work is supported by the Russian Ministry of Science and Education under contract №14.Z50.31.0024.

References

[1] ATLAS Collaboration, 2008 The ATLAS Experiment at the CERN Large Hadron Collider // Journal of Instrumentation 3 S08003
URL: http://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08003/pdf [accessed on: 26.10.2018]

[2] Trigger whiteboard [Online]. Available:
https://atlasop.cern.ch/twiki/bin/view/Main/TriggerWhiteBoard [accessed on: 17.10.2018]

Data Quality whiteboard [Online]. https://atlasop.cern.ch/twiki/bin/view/Main/DQWhiteBoard [accessed on: 17.10.2018]

MC production campaign 'mc15' overview [Online].
https://twiki.cern.ch/twiki/bin/view/AtlasProtected/AtlasProductionGroupMC15FullSummary [accessed on: 17.10.2018]

DDM space tables [Online]. http://adc-ddm-mon.cern.ch/ddmusr01/ [accessed on: 17.10.2018]

[3] K. Cranmer et al., 2015 Analysis Preservation in ATLAS // Journal of Physics: Conference Series 664 032013 URL: http://iopscience.iop.org/issue/1742-6596/664/3 [accessed on: 17.10.2018]

[4] ATLAS Metadata Interface [Online]. Available: https://ami.in2p3.fr/ [accessed on: 19.10.2018]

[5] GLANCE project [Online]. Available: https://glance.cern.ch/ [accessed on: 17.10.2018]

[6] M. Grigoryeva, M. Golosova, A. Klimentov, T. Wenaus, Data Knowledge Base for HENP Scientific Collaborations // Journal of Physics: Conference Series – Proceedings of the 18th International Workshop on Advanced Computing and Analysis Techniques in Physics Research

[7] M. Grigoryeva, V. Auliv, A. Klimentov, M. Gubin, 2015 Knowledge Base of a Scientific Experiment // Open systems. DBMS. № 4. pp.14–17.
URL: https://www.osp.ru/os/2016/04/13050998/ [accessed on: 17.10.2018] (in Russian)

[8] M. Golosova, M. Grigorieva, A. Klimentov, E. Ryabinkin, G. Dimitrov and M. Potekhin, 2015 Studies of Big Data metadata segmentation between relational and non-relational databases // Journal of Physics: Conference Series 664 042023
URL: http://iopscience.iop.org/article/10.1088/1742-6596/664/4/042023/pdf [accessed on: 17.10.2018]

[9] ProdTask interface [Online]. Available: https://prodtask-dev.cern.ch/prodtask/ [accessed on: 26.10.2018]

[10] BigPanda monitor [Online]. Available: https://bigpanda.cern.ch/ [accessed on: 26.10.2018]

[11] Elasticsearch [Online]. Available: https://www.elastic.co/products/elasticsearch [accessed on: 17.10.2018]

[12] ATLAS Dataset Nomenclature [Online]. Available: https://dune.bnl.gov/w/images/9/9e/Gen-int- 2007-001_%28NOMENCLATURE%29.pdf [accesses on: 17.10.2018]

[13] DKB project repository [Online]. Available: https://github.com/PanDAWMS/dkb/ [accessed on: 17.10.2018]
DKB Elasticsearch data scheme [Online]. Available: https://github.com/PanDAWMS/dkb/blob/master/Utils/Elasticsearch/mapping/tasks.mapping [accessed on: 17.10.2018]

[14] Rucio DDM [Online]. Available: https://rucio.cern.ch/ [accessed on: 26.10.2018]

[15] Apache Kafka [Online]. Available: https://kafka.apache.org/ [accessed on: 26.10.2018]

[16] CERN Document Server [Online]. Available: https://cds.cern.ch/ [accessed on: 17.10.2018]

[17] F. H. Barreiro, M. Borodin, K. De, D. Golubkov, A. Klimentov, T. Maeno, R. Mashinistov, S. Padolski, T. Wenaus, on behalf of the ATLAS Collaboration, 2017 The ATLAS Production System Evolution: New Data Processing and Analysis Paradigm for the LHC Run2 and High-Luminosity // Journal of Physics: Conference Series 898 052016
URL: http://iopscience.iop.org/article/10.1088/1742-6596/898/5/052016/pdf [accessed on: 26.10.2018]