

The new approach to monitor the workflow management system ProdSys2/PanDA of the ATLAS experiment at LHC by using methods and techniques of visual analytics

T.Galkin^{A,1}, M.Grigoryeva^{B,D,2}, A.Klimentov^{B,C,3}, T.Korchuganova^{D,4}, I.Milman^{A,5}, S.Padolski^{C,6}, V.Pilyugin^{A,7}, D.Popov^{A,8}, M.Titov^{B,9}

^A National Research Nuclear University "MEPhI", Russia

^B National Research Center "Kurchatov Institute", Russia

^C Brookhaven National Laboratory, USA

^D National Research Tomsk Polytechnic University, Russia

¹ ORCID: 0000-0003-2859-6275, TPGalkin@mephi.ru

² ORCID: 0000-0002-8851-2187, magsend@gmail.com

³ ORCID: 0000-0003-2748-4829, Alexei.Klimentov@cern.ch

⁴ ORCID: 0000-0001-5792-8182, tatiana.korchuganova@cern.ch

⁵ ORCID: 0000-0001-9705-9401, igalush@gmail.com

⁶ ORCID: 0000-0002-6795-7670, spadolski@bnl.gov

⁷ ORCID: 0000-0001-8648-1690, VVPilyugin@mephi.ru

⁸ ORCID: 0000-0002-3333-749X, DDPopov@mephi.ru

⁹ ORCID: 0000-0003-2357-7382, mikhail.titov@cern.ch

Abstract

The paper presents the pilot project "Tools and methods of visual analytics as part of the workload management system ProdSys2/PanDA of the ATLAS experiment at the Large Hadron Collider." The project is aimed at expanding the functionality of the existing ATLAS monitoring system using a visual analytics approach to analyze large volumes of multidimensional data of computing tasks and jobs. The functioning of ATLAS workflow management system (WMS) is associated with the processing of multi-petabytes and exabytes data volumes. In this case, the new challenges arise that require the use of multi-level interactive visualization tools to analyze the correlations between individual datasets and their representations. The article contains the description of the proposed approaches to visual analysis of multidimensional data, as well as the definition of areas of use of visual analytics applications for data processing in the ATLAS experiment. The tasks for approbation of the recommended methods using the example of the analysis of statistical data of ATLAS WMS are determined. The final part of the article is devoted to the organizational component of the pilot project.

Keywords: visual analytics, workflow management system, big data, LHC.

1. Introduction

The pilot project "Tools and methods of visual analytics as part of the workload management system (WMS) ProdSys2/PanDA of the ATLAS experiment at the Large Hadron Collider (LHC)" is aimed at developing state-of-the-art visualization and analytical approaches, and tools for monitoring distributed WMS using the example of data processing and analysis system of the ATLAS experiment [1] at LHC [2] (CERN, Switzerland). Initial motivation of the project is associated with

the experiments at LHC, but both quantitative and qualitative requirements are common for many experiments in high energy and nuclear physics (HENP). Therefore this project is of interest to a wide range of scientific groups, as the discussion under of the 26th Symposium on Nuclear Electronics and Computing (NEC2017) has shown [3].

Internal information (such as, description of format and every stage of processing of data from physical experiments) of distributed data processing systems should be processed, analyzed and pre-

sented to users of the system in a suitable and compact form. For this purpose, specialized monitoring systems are being developed. The paper [4], also published in this issue of the journal, describes the monitoring system currently existing in the ATLAS experiment - BigPanDA. The system includes following possibilities of data visualization: interactive interfaces, parametric tables and various charts (histograms, bar and pie charts, simple two-dimensional graphics). Until now, the requirements for the monitoring system have been limited to the use of basic visual analysis methods for a limited class of tasks, and data dimensionality limited to three dimensions. However, the constant increase in the amount of processed data and the complexity of the computing infrastructure of WMS of the ATLAS experiment, as will be shown below, produces new challenges related to the visual analysis of large volumes of multidimensional data.

Under this project it is proposed to apply a fundamentally new approach to monitoring the operation of complex distributed systems, and to change the "classical" monitoring of WMS (particularly, in the field of HENP) using visual analysis methods [5,6] to improve the quality of metrics for data state definition. The application of these methods will allow to remove the restriction on the number of dimensions of the analyzed data, providing the generation of multidimensional geometric interpretations for visual analysis. As a result, the functionality of the monitoring system will be expanded by the ability to detect explicit correlations between different data objects and the use of multi-level interactive interfaces. In addition, the joint application of visual analytics and "machine learning" methods will increase the level of automation in the functioning of data processing and analysis systems primarily in the field of particle physics.

2. Issues of management of extremely large data

The scientific research cycle on modern installations can last for decades (for ex-

ample, scientific communities on the LHC were established more than 25 years ago, the data collection was started 9 years ago and will last at least another 10 years). During the lifetime of the experiments, the installations and detectors are modernizing, information technologies are qualitatively changing. At the same time, the software and hardware infrastructure as well as the data processing model (computing model) are evolving: the number of data centers is constantly increasing (new architectural solutions are emerging), the scenarios for launching and performing analysis, data processing and modeling tasks are changing, software versions are being updated, the technologies of storage and access to data and metadata becomes different. Under a constantly evolving and complex computing infrastructure as well as a simultaneous growth of information flow, it becomes difficult to control the operating of data management systems, organize data processing and simulation of the experiment, predict and timely detect possible anomalies in the functioning of individual hardware and software components of the computing infrastructure and data processing systems. To solve these problems specialized analytical tools are being developed in modern experiments. A separate crucial task is to search for patterns or analysis of anomalies in the operating of complex distributed systems, correlations between the actions of the operator's service and the behavior of the system, prediction of system performance in case of software changes.

The second generation of the ProdSys2 (Production System) [7] of the ATLAS experiment, in conjunction with the workload management system PanDA [8] (Production and Distributed Analysis system), that is a complex set of hardware and software tools for organizing, planning, starting and executing computing tasks and jobs (Figure 1). ProdSys2/PanDA is responsible for all stages of data processing, analysis and modeling, including simulating of physical processes and functioning of the detector using the Monte Carlo method, (re)processing of physical data, performing highly specialized tasks (e.g., setting a

high-level trigger or a software quality assurance task). Using the ProdSys2/PanDA software, the ATLAS scientific community, individual physical groups and scientists have access to hundreds of WLCG [9] (Worldwide LHC Computing Grid) computing centers, supercomputers, cloud com-

puting resources and university clusters. Characteristics of the system can be reflected by the following indicators: implementation of more than a million computing tasks per day at 200+ computer centers by thousands of users utilizing more than 300,000 nodes.

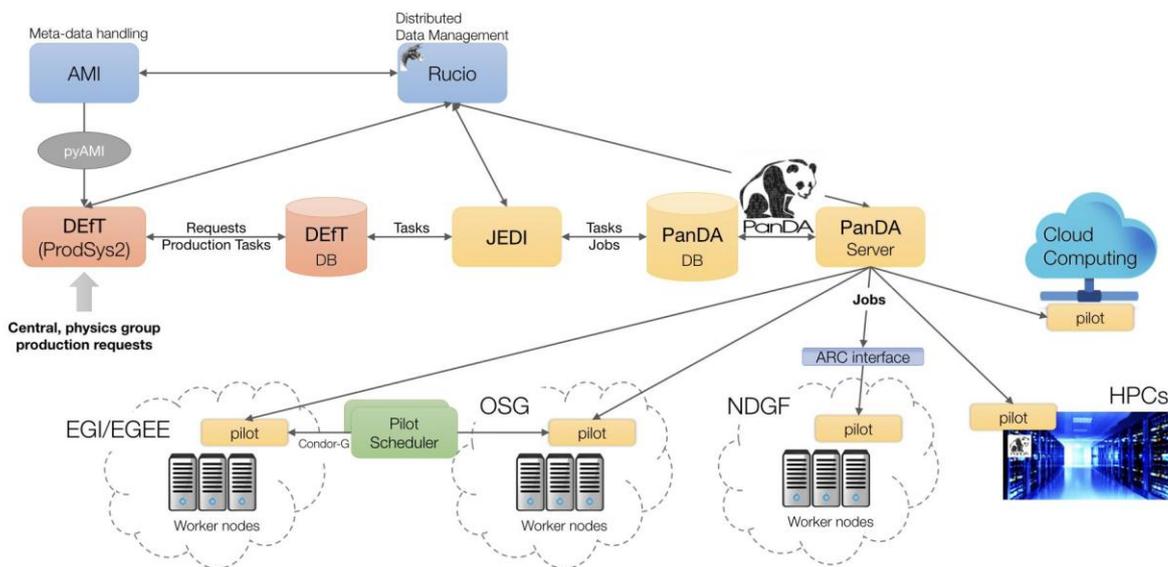


Fig.1. Data processing workflow at the ATLAS experiment

In a distributed system, there is always a competition between different threads of computing tasks. For example, in the period preceding the main physical conferences, the number of data analysis tasks distinctly increases (from Figure 2 it follows that in the ATLAS experiment there was a sharp increase in the number of computing jobs performed in certain months). Therefore, when starting computing jobs, there can be significant delays due to the lack of a free computing resource. As a rule, the end user is interested not so much in the process of computing jobs execution as in the ability to predict the data processing (or analysis) completion time and get a scientific result in a predetermined time period (it can be hours for analysis tasks, or weeks for data processing). The tasks flow itself has several phases of execution (e.g., modeling, digitization, reconstruction, creation of objects for physical analysis, physical analysis it-

self) and each stage (phase) can be performed on geographically distributed computing centers, which includes the transfer of initial input data between computing centers and can affect the total execution time of the entire computing tasks flow. Possible hardware malfunctions may require redistribution of tasks between computing centers, that also provides additional ambiguity in the prediction of the data processing metrics. At the moment there is no central portal for monitoring, the operator is forced to view data transfer schedules, task performing schedules, tables with information about the computing centers and individual components functioning. The creation of a single portal and the ability to visualize the operation of systems will allow to optimize the utilization of a computing resource, significantly automate and simplify the workflow of data processing, and thereby accelerate the delivering of scientific results.

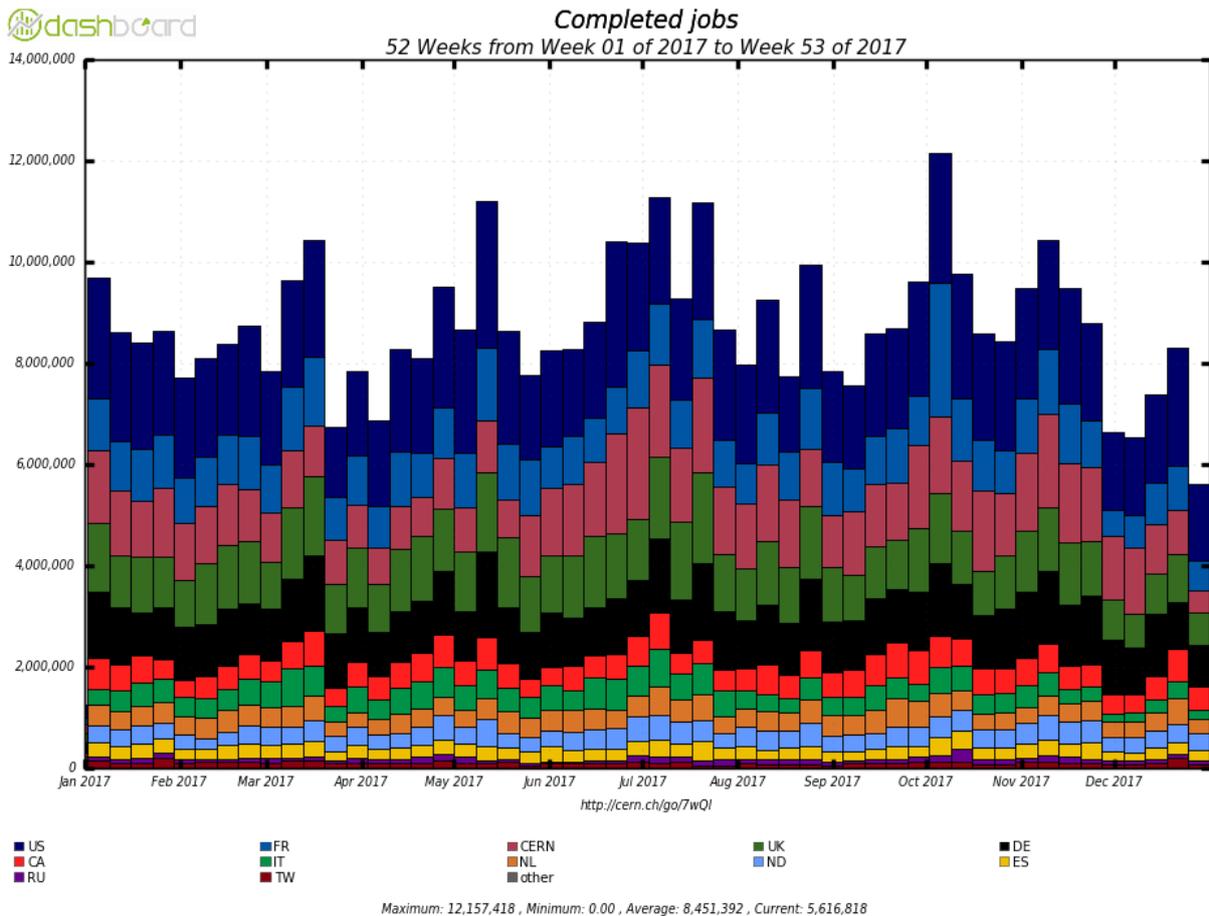


Fig.2. Number of completed computing jobs grouped by computing “regions” during 2017 (average per week)

Information that was accumulated over the entire period of operation of WMS of the ATLAS experiment (this is more than 14 years) contains records of the progress of execution of more than 10 millions of computing tasks and about 3,000 millions of computing jobs. Based on such statistics it allows to use "machine learning" methods to make analytical calculations and to forecast the software functioning. The work related to ProdSys2/PanDA, in which the development and usage of "machine learning" methods for the analysis of processing data [10,11] were started, showed that the calculated metrics based on the predictive analytics allow to increase the efficiency of the processing of real and simulated data, providing more thorough planning of the analysis process (determined by individual users or a group of users), predicting possible failure or abnormal behavior of the system (produced by agents of control services).

3. Visual analytics approach

In the modern world, problems related to processing and analysis of multidimensional data are among the most urgent tasks. Many different methods and hardware-software tools are developed to solve these tasks, that also includes both automatic and interactive solutions. It is worth noting that data visualization can be used within these methods. Currently, the approach of visual analytics is widely used. This approach was preceded by solutions for solving tasks of multidimensional data analysis by a variety of visual methods.

A review of the literature describing specific applications using visual methods makes it possible to assert that, in reality, interactive systems with multidimensional data are often given less importance than systems for displaying the results of applying data analysis methods. Examples include systems such as the situational

awareness system AdAware [12], the visual analysis system in aircraft engineering [13], the SAS Visual Analytics software package [14], designed to process and analyze large amounts of business data. Experience has shown that classical methods of parallel coordinates, Andrews curves, Chernoff faces and other similar mnemonic graphical representations are widely used for visual representation of multidimensional data. All these visual methods are based on the fact that the analyzed tuples of numerical data are interpreted as the values of the parameters of such mnemonic graphical maps. Examples of such mappings are shown in Figures 3,4.

It should be noted that all systems using the above visual methods are essentially configured to internal processing of multidimensional data and presentation of it to analysts in a convenient form. They do not provide an opportunity to work with the analyzed data directly and their multivariate geometric interpretations using corresponding visual representations that are natural for human. The approach of visual analytics involves solving data analysis tasks, and, in particular, tasks of multidimensional data analysis, using a conducive interactive visual interface.

One of the most common forms of visual analytics is the solution for tasks of multidimensional data analysis by the visualization method [17]. The solution of the task of the initial data analysis by the visualization method consists of the sequential solution of the following two subtasks (Figure 5).

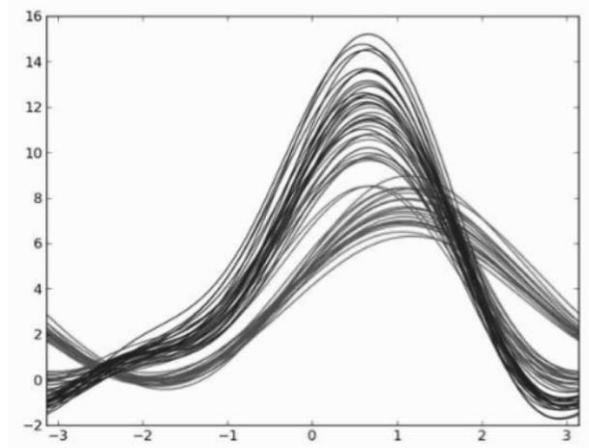


Fig.3. Fisher's Iris data set presented in the form of the Andrews curves [15]

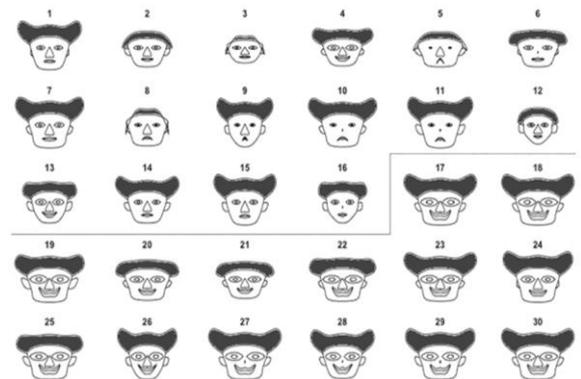


Fig.4. Chernoff faces for medical data [16]

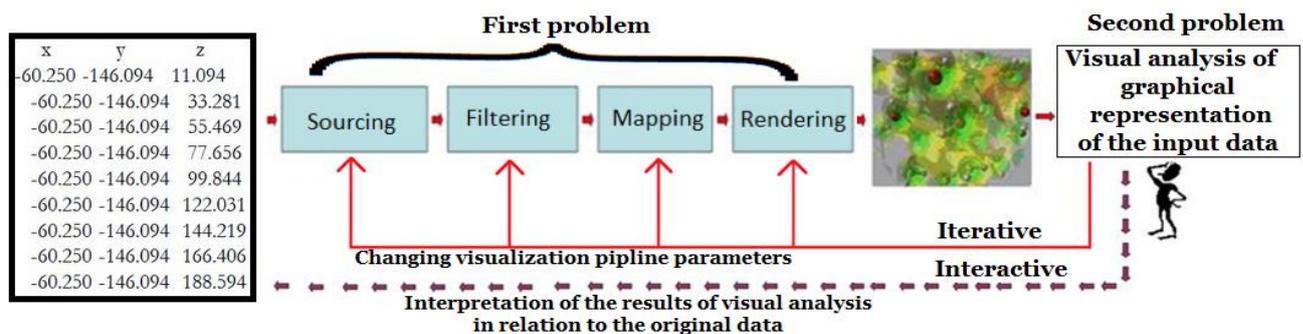


Fig.5. Data analysis using the visualization method

The first subtask is to get a representation of the analyzed data in the form of a certain graphic image (the visualization of the original data), which is solved using a computer. The resulting graphic images serve as a natural and convenient means to represent the spatial interpretation of the initial data to the person (analyst). Spatial interpretation is one or more spatial objects (i.e., the spatial scene), which are set in compliance to the analyzed data. The second subtask, which is no less important, is the visual analysis of the graphical representation of the analyzed data obtained as a result of solving the first subtask, while the analysis results are interpreted with respect to the original data, and which is solved directly by the analyst. The spatial scene is visually analyzed using the enormous potential possibilities of the analyst's spatial and conception thinking during the

analysis. As a result of solving this problem, the analyst makes some judgments about the spatial scene. Thus, judgments about the object under consideration are formulated. The process of visual analysis of the graphic image is not strictly formalized. The efficiency of visual analysis is determined by the experience of the person who carries out this image analysis and his propensity for spatial and conception thinking. Looking at the resulting image, a person is able to solve 3 main tasks: analysis of the shape of spatial objects, analysis of their relative positions and analysis of graphic attributes of spatial objects.

Under this project, it is proposed to use the multidimensional geometric modeling of the initial data, which are considered as multidimensional tabular data about computing jobs (Table 1)

Table 1. Representation of computing jobs (multidimensional tabular data)

	Parameter 1	Parameter 2	...	Parameter n
Job 1	x_1^1	x_1^2	...	x_1^n
...	
Job i	x_i^1	x_i^2	...	x_i^n
...
Job m	x_m^1	x_m^2	...	x_m^n

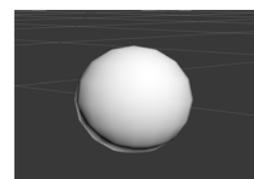
A geometric interpretation is carried out to solve the defined task. Rows of the table correspond to multidimensional points in the space E_n , $p_i = (p_i^1, p_i^2, \dots, p_i^n) \in E_n$, and the values of the computing job parameters are the coordinates of multidimensional points. It is suggested to interpret the difference in parameter strings as the Euclidean distance between the points of this multidimensional space (the longer the distance is, then the lines are more different).

With this interpretation, the analysis of the distance between points of the N-dimensional space is assigned to the analysis task of similarities and differences in records of computing jobs.

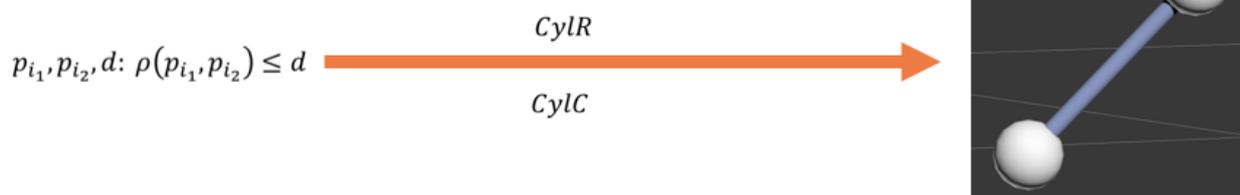
It is proposed to use a visual presentation of points in the N-dimensional space to analyze the distance between these points. At the beginning, the original set of points is projected onto one of the three-dimensional spaces. Wherein:

- The multidimensional point p_i is projected into the sphere S_i .

$$p_i = (p_i^1, p_i^2, \dots, p_i^n) \xrightarrow[p_{iC}]{p_i^{l_1}, p_i^{l_2}, p_i^{l_3}, SphR}$$



- If the distance between the points of the N-dimensional space p_1 and p_2 is less than the threshold distance d given by the analyst in an interactive mode, then a cylinder is constructed to connect the spheres S_1 and S_2 .
- The color of the cylinder simulates the distance between the points p_1 and p_2 from red (small distance) to blue (long distance).



Then, the graphic projection of spheres and cylinders onto the picture plane is performed, and followed by its corresponding visual analysis. The resulting set of spheres and cylinders forms a spatial scene with a given geometry and optical (color) characteristics.

Thus, visual analysis of the spatial scene will allow to judge the distance between the original multidimensional points. In the process of solving the analysis task, it is proposed to set the initial large value of d , and then reduce it and select subsets of multidimensional points, depending on the resulting image in the picture plane. It should be noted that under this approach, the analyst in the process of data analysis does not passively contemplate the spatial scene, but has the possibility of interactive engagement with it.

4. Areas of use of visual analytics

4.1. Discovery of relations (and influence) between parameters of data objects

Interpretation of the parameters of data objects and corresponding metrics in the N-dimensional space (the dimension is determined by the number of parameters to be investigated) to evaluate the mutual influence and determine their correlations.

4.2. Automatic anomaly detection while performing tasks of processing, analysis and modeling

Interpretation of data in the N-dimensional space (dimensionality is determined by the number of parameters to be investigated) and projection into 3-dimensional space will allow clustering of objects (by specified parameters and metrics) in order to detect objects with non typical set of values of specified parameters (i.e., anomalous objects). In the case of distributed workflow management system, such data objects are computing tasks and jobs (computing task consists of a set of computing jobs) that describe the processing of data in the corresponding specialized systems (ProdSys2 and PanDA). The collection of the necessary information about data processing (object parameters and metrics) is performed during the definition of computing tasks and jobs, and their subsequent execution (it is possible to take into account parameters and metrics which describe the current state of the computing environment and the computing resources used).

4.3. Determination of the reasons for the inefficiency of processes for data processing

Determination of the sequence of stages in which there was a failure, a delay or an error during the data processing, and the

identification of a set of values of the specified parameters in the initial stage that caused the failure (based on the methods developed in section 4.1).

4.4. Data popularity estimation for dynamic data management

The popularity of data is determined by the number of accesses of analysis tasks to datasets (i.e., data objects) and the number of requests for additional dataset replicas/copies (at different computer centers). When the number of accesses to datasets is increased, then the additional data replicas should be created automatically, and that replicas should be "tied" geographically to the computer center with the highest demand.

The studies carried out based on data from the monitoring system showed that the "popularity" of data is dramatically reduced after about 45 days, which allows to make a decision to delete replicas of not used datasets from disks and to transfer them to tape.

Visual analytics methods will allow to cluster the data by geographical location of storage (computing center) and their demand for analysis depending on time. This will optimize the requirements for creation/deletion of additional/redundant data replicas.

5. Approbation of the proposed approach

5.1. Visual analytics of statistical data about computing jobs of the ProdSys2/PanDA system

5.1.1. Problem statement

Development of visualization tools and methodology of using them for cluster analysis of tuples of parameters of computing jobs of the PanDA system. Visual analysis will allow to identify similar computing jobs, as well as to reveal anomalous computing jobs, and at the same time determining parameters that caused this anomaly.

Development of visualization tools and methodology of using them to analyze execution durations of computing jobs of the PanDA system. Visual analysis will allow to determine the influence of a certain set of parameters (i.e., basic set of parameters, which can be extended in the future) on the execution time of computing jobs, and the identification of a set/range of values of individual parameters indicating an increase in the execution time of computing jobs (i.e., the deviation from the average time, it affects tasks with execution time exceeding 3σ , it is assumed that this is about 7.5% of the number of all tasks).

Development of visualization tools and methodology of using them to analyze the popularity of data in relation to time. Visual analysis will reveal the increase/decrease in the number of accesses to data as a function of time, thus providing a convenient visual method for decision making process of the dynamic management of the data replicas distribution.

5.1.2. Expected results

Creation of projection graphic images of the N-dimensional geometric interpretations of parameters of computing jobs, 2D histogram representation of the number of computing jobs (Y axis) grouped by their execution durations (X axis), and identification of a group of computing jobs with "increased" execution duration (more than average).

5.1.3. Data description

To solve the defined research task, key parameters of computing jobs are defined:

- Duration of the computing job execution ($\langle \text{duration} \rangle = \text{endtime} - \text{starttime}$)
- Name of the job flow-group (gShare)
- Computing center for the job processing (nucleus)
- Number of events to be processed (nEvents)
- Additional set of parameters (extended set of parameters for the basic set):
 - The name of the analysis/processing stage (pro-

cessingType); the amount of input data for the job (inputFileBytes); the type of input data (inputFileType); the amount of output data of the job (outputFileBytes); the initial priority of the job (assignedPriority); CPU time to process one event (cpuTimePerEvent); the hardware architecture on which calculations are performed for the job (cmtConfig); the number of cores (actualCoreCount); software release (atlasRelease); CPU efficiency per core (CPU eff per core); the average size of memory pages allocated to the process by the operating system and currently located in RAM (avgRSS); the average fraction of shared memory used by the CPU (avgPSS); the average size of the allocated virtual memory (avgVMEM); the maximum size of memory pages allocated to CPU by the operating system (maxRSS); the maximum share of the total memory used by CPU (maxPSS); the maximum size of the allocated virtual memory (maxVMEM)

- Indication parameters:
 - Step of restarting the job (attemptNr); error codes (brokerageErrorCode, ddmErrorCode, exeErrorCode, jobDispatcherErrorCode, pilotErrorCode, supErrorcode, taskBufferErrorcode)

The format of the raw data representation:

- Combination of job parameters into groups
- The representation of input data in the form of matrices corresponding to a group of parameters, where rows correspond to job records, and columns correspond to the parameters of a certain group:
 - $D_{n \times 1}$ - matrix with jobs' durations, where n - the number of jobs;

- $P_{n \times 3}$ - matrix with the basic set of job parameters (gShare, nucleus, nEvents);
- $E_{n \times 16}$ - matrix with the additional/extra set of parameters;
- $I_{n \times 8}$ - matrix with the set of indication job parameters (attemptNr, errorCodes).

Data source:

- Infrastructure Elasticsearch at the University of Chicago [18]
- Indices "jobs_archive_*"
 - Search conditions
 - Acceptable statuses of jobs: jobStatus IN ("finished", "failed");
 - Source of jobs: prodSourceLabel = 'managed'
 - Type of processing data and the stage of processing: REGEXP_LIKE (jobName, "^mc(.*\.)\{3\}simul\..*")

5.2. Visual analytics of statistical data about execution of computing tasks of the ProdSys2/PanDA system

This research task implies an extension of the task/problem in section 5.1 applied to computing tasks of ProdSys2 and the solution proceeds from the results of the research task given in 5.1, implies a similar approach, but taking into account the specifics of the data objects that are under consideration - computing tasks, on the basis of which sets of computing jobs are formed.

6. Primary stages of project implementation

Approbation of the approach of visual analytics using the example of applying the cluster analysis of tuples of the computing jobs parameters of PanDA system and estimating the distribution of computing jobs' durations.

Expansion of the developed approach with respect to computing tasks of ProdSys2.

Integration of the developed prototypes of visualization and analytical tools into the monitoring infrastructure of the ProdSys2/PanDA system.

Evaluation of the modification of the existing ProdSys2/PanDA monitoring using the visual analytics approach.

7. Expected project results

As a result of the project, a visual analytics system will be developed to monitor workflow management systems. The developed system will be an extended analytical service of the existing ATLAS monitoring system. By means of the developed system, the monitoring functionality will be significantly expanded, allowing to simulate, predict the further course and state of the experiment. Visual analytics will form the basis of a decision support system and strategic planning.

Cooperation with the ATLAS experiment at LHC, availability of access to experimental data and demonstration of the created solution and prototype on an existing data processing system, will provide a unique testing ground for developing analytical research technologies and application of visual analytics methods and will allow this project to be among the world's most important developments of the given area.

The results of the project will be in demand for the creation of the software for the NICA collider (JINR, Dubna), for the high-luminosity LHC (HL-LHC), and for visualization of scientific information on mega-facilities such as XFEL and FAIR.

8. Primary participants

The ATLAS collaboration in joint with Russian research centers and universities participate in this pilot project.

List of research centers and universities:

- National Research Nuclear University "MEPhI"
 - Laboratory of Scientific Visualization
 - Department of Analysis of Competitive systems

- MEPhI group in the ATLAS experiment
- National Research Center "Kurchatov Institute"
 - Laboratory of Big Data Technologies
- National Research Tomsk Polytechnic University
- Joint Institute for Nuclear Research
 - Laboratory of Information Technologies
- Brookhaven National Laboratory
- University of Iowa
- University of Chicago
- University of Texas at Arlington
- European Organization for Nuclear Research (CERN)

9. Work with students and teaching activities

One of the objectives of the project is to work with students, including the preparation of bachelors, masters and graduate students who manage to use advanced tools for scientific visualization and work with data from a physical experiment. The courses "Visual Analytics" and "Scientific Visualization" are taught at the NRNU MEPhI, on the basis of the results of the project, special courses for graduate students with major in particle and nuclear physics, and system engineering will be created. In addition, the University of Dubna and the Institute of Cybernetics of TPU expressed interest in creating joint courses on the subject of the project (the University of Dubna has created a course for training/educating specialists for work at the NICA collider. TPU actively participates in the scientific program in the field of particle physics: the COMPASS experiment on a super-proton synchrotron (SPS, CERN), and ATLAS and CMS experiments at LHC).

10. Information support

Information support for the project is provided by the journal "Scientific Visualization" [19], as well as by the informational portals of the ATLAS experiment at CERN

[20], the Laboratory of Big Data Technologies NRC KI [21] and LIT JINR [22].

References

1. The ATLAS Collaboration, "The ATLAS Experiment at the CERN Large Hadron Collider", *Journal of Instrumentation*, vol. 3, S08003, 2008.
2. LHC - Large Hadron Collider, <http://lhc.web.cern.ch/lhc>
3. 26th International Symposium on Nuclear Electronics & Computing - NEC'2017, <http://indico.jinr.ru/conferenceDisplay.py?confId=151>
4. S.Padolski, T.Korchuganova, T.Wenaus, M.Grigorieva, A.Alexeev, M.Titov, A.Klimentov, "Data visualization and representation in ATLAS BigPanDA monitoring", *Scientific Visualization*, 2018.
5. J.Thomas, K.Cook, "Illuminating the Path: The Research and Development Agenda for Visual Analytics", *IEEE Computer Society*, 2005.
6. D.Popov, I.Milman, V.Pilyugin, A.Pasko, "Visual Analytics of Multi-dimensional Dynamic Data with a Financial Case Study", *Data Analytics and Management in Data Intensive Domains*, Springer International Publishing, pp. 237--247, 2017.
7. M.Borodin, K.De, J.Garcia Navarro, D.Golubkov, A.Klimentov, T.Maeno, A.Vaniachine, "Scaling up ATLAS production system for the LHC Run 2 and beyond : project ProdSys2", *Journal of Physics: Conference Series*, vol. 664, no. 6, 2015.
8. A.Klimentov et al., "Migration of ATLAS PanDA to CERN", *Journal of Physics: Conference Series*, vol. 219, no. 6, 2010.
9. WLCG - Worldwide LHC Computing Grid, <http://wlcg.web.cern.ch>
10. M.Titov, G.Zaruba, A.Klimentov, and K.De, "A probabilistic analysis of data popularity in ATLAS data caching", *Journal of Physics: Conference Series*, vol. 396, no. 3, 2012.
11. M.Titov, M.Gubin, A.Klimentov, F.Barreiro, M.Borodin, D.Golubkov, "Predictive analytics as an essential mechanism for situational awareness at the ATLAS Production System", *The 26th International Symposium on Nuclear Electronics and Computing (NEC), CEUR Workshop Proceedings*, vol. 2023, pp. 61--67, 2017.
12. Y.Livnat, J.Agutter, S.Moon, S.Foresti, "Visual correlation for situational awareness", *IEEE Symposium on Information Visualization (INFOVIS)*, pp. 95--102, 2005.
13. D.Mavris, O.Pinon, D.Fullmer, "Systems design and modeling: A visual analytics approach", *Proceedings of the 27th International Congress of the Aeronautical Sciences (ICAS)*, 2010.
14. SAS the power to know, [Online]. Available: http://www.sas.com/en_us/home.html [accessed on 15.03.2018].
15. K.Sharopin et al., *Vizualizacija medicinskih danyh na baze paketa NovoSpark* ["Vizualization of the Medical Data on the basis of package NovoSpark"], *Izvestiya SFedU. Engineering Sciences*, vol. 109, pp. 242--249, 2010. [In Russian]
16. J.Woollen, "A Visual Approach to Improving the Experience of Health Information for Vulnerable Individuals", *PhD Thesis, Columbia University Academic Commons*, 2018.
17. V.Pilyugin et al., "Scientific visualization as method of scientific data analysis", *Scientific Visualization*, vol. 4, pp. 56--70, 2012.
18. <http://atlas-kibana.mwt2.org:5601/app/kibana>
19. <http://sv-journal.org>
20. <http://atlas.cern>
21. <http://bigdatalab.nrcki.ru>
22. <http://lit.jinr.ru>