2023, **24** (1), 1–9. doi 10.26089/NumMet.v24r101

doi 10.26089/NumMet.v24r101

Bottlenecks in organizing the workflows of large HPC centers

Dmitry A. Nikitenko
Lomonosov Moscow State University, Research Computing Center, Moscow, Russia
ORCID: 0000-0002-2864-7995, e-mail: dan@parallel.ru

Abstract: Effective output from data centers are determined by many complementary factors. Often, attention is paid to only a few, at first glance, the most significant of them. For example, this is the efficiency of the scheduler, the efficiency of resource utilization by user tasks. At the same time, a more general view of the problem is often missed: the level at which the interconnection of work processes in the HPC center is determined, the organization of effective work as a whole. Omissions at this stage can negate any subtle optimizations at a low level. This paper provides a scheme for describing workflows in the supercomputer center and analyzes the experience of large HPC facilities in identifying the bottlenecks in this chain. A software implementation option that gives the possibility of optimizing the organization of work at all stages is also proposed in the form of a support system for the functioning of the HPC site.

Keywords: supercomputing, provision of computing resources, use of computing resources, workflows at supercomputer center, shared research facilities, provision of computing services.

Acknowledgements: The reported study was funded by RFBR, project number 20–07–00864. The research is carried out using the equipment of the shared research facilities of HPC computing resources at Lomonosov Moscow State University.

For citation: D. A. Nikitenko, “Bottlenecks in organizing the workflows of large HPC centers,” Numerical Methods and Programming. **24** (1), 1–9 (2023). doi 10.26089/NumMet.v24r101.

Узкие места в организации рабочих процессов больших суперкомпьютерных центров

Д. А. Никитенко
Московский государственный университет имени М.В. Ломоносова, Научно-исследовательский вычислительный центр, Москва, Российская Федерация
ORCID: 0000-0002-2864-7995, e-mail: dan@parallel.ru

Аннотация: Эффективная отдача от вычислительных центров определяется множеством взаимодополняющих факторов. Зачастую внимание уделяется только нескольким, на первый взгляд наиболее значимым из них. Например, это эффективность работы планировщика, эффективность использования ресурсов пользовательскими задачами. При этом часто упускается более общий взгляд на проблему: уровень, на котором определяется взаимосвязь рабочих процессов в СКЦ, организация эффективной работы в целом. Упущения на этом этапе способны обесценить любые тонкие оптимизации на низком уровне. В данной работе приводится схема описания рабочих процессов в СКЦ, анализируется опыт крупных СКЦ по выделению наиболее узких мест в этой цепочке, предлагается вариант программной реализации, дающей возможность оптимизации организации работ на всех этапах, в виде системы поддержки функционирования СКЦ.

Ключевые слова: суперкомпьютинг, предоставление вычислительных ресурсов, использование вычислительных ресурсов, организация работы СКЦ, вычислительные центры коллективного пользования, предоставление вычислительных услуг.

Благодарности: Исследование выполнено при финансовой поддержке РФФИ (грант 20–07–00864). Работа выполнена с использованием оборудования Центра коллективного пользования сверхвысокопроизводительными вычислительными ресурсами МГУ имени М. В. Ломоносова.

Для цитирования: Никитенко Д.А. Узкие места в организации рабочих процессов больших суперкомпьютерных центров // Вычислительные методы и программирование. 2023. **24**, № 1. 1–9. doi 10.26089/NumMet.v24r101.

1. Introduction. The questions of increasing the efficiency of supercomputer computing systems have always been acute due to the high cost and scarcity of this type of resources. The choice of target software and hardware architectures, based on the planned type of tasks to be solved, fine-tuning of the environment, optimizing the applications themselves — there are a lot of scientific and practical events and articles are devoted to all this.

Various approaches to assessing the quality of compute services are discussed in [1]. Meanwhile, a higher level is much less affected — the level of organization of the entire chain of work processes from the very idea of using supercomputer resources to obtaining the required result. Even in the State rubricator of scientific and technical information classification there was a special block dedicated to this direction. At the moment, this kind of work receives somewhat less publicity, but not because it is not relevant, rather since its target audience is significantly narrower and consists mainly of holders and administrators of computer systems, moreover, on a fairly large scale.

Most of the related contemporary works discuss the questions of managing and maintenance of the specified computing facilities [2] or [3, 4], giving some best practices, but with almost no suggestions, how to solve the discussed problems in general.

Within the framework of the project, based on the results of which this work was written, the issues of identifying the main routines and stages in the operation of a supercomputing center were studied. The main purpose of this study is to identify bottlenecks at these stages and an appropriate choice of possible approaches to eliminate them when organizing work to provide computing resources.

In the first substantive section of the paper, the state of the subject area is considered, namely an analysis of the survey conducted among the holders and administrators of large supercomputer centers is given. The following is a description of the proposed model of the functioning of the HPC center, after which the software implementation corresponding to it is addressed.

2. Background. To assess the current vision of system owners and system administrators of the most critical moments in the organization of provision of resources, a list of questions was compiled, to which 12 supercomputing centers, both from Russia and Europe, kindly agreed to answer. The survey included questions about the scale and profile of the computing center, and the features of its operation and support. Among other things, respondents were asked to indicate the degree of significance of certain stages and routines in the overall chain.

Questions on access to HPC resources:

1. What computing systems do you have at HPC center?
2. Are there various compute partitions used?
3. If used, what is the principle of division?
4. What is the size of the team supporting the HPC center?
5. What is the vendor's role in maintaining the systems?
6. Do systems shutdown during the year?
7. How regularly is the maintenance done (the works which cause partial nodes unavailability)?

8. Approximate number of user accounts
9. Approximate number of users (researchers)
10. What classification of users is used? (meaning roles: administrator, user, project manager, etc.)
11. What organizations are your users from? (approximate ratio of internal and external)
12. Is the provision of resources allowed on a commercial basis?
13. What steps should a user take to gain access?
14. If a project organization is practiced (resources are granted for a specific research project), what steps should the project participants take?
15. Whom are the computing resources allocated to? (account, group of accounts, individual, team, etc.)
16. What types of quotas and limits are used? (limiting the total amount of resources, specific resource consumption, the number of simultaneously running tasks, the number of tasks in the queue, the duration of the task, the size of the task, etc.)
17. How is the control over the consumption of computing resources carried out?
18. Should users provide reports on resource usage?
19. How is the helpdesk organized for users?
20. What software is used to organize the workflow (registration of requests for the allocation of resources, register of projects, users, etc.)
21. What routines are covered by the supporting software (license control, warranty control, initial provisioning, resource monitoring, etc.)?
22. How do users see the availability of resources?
23. Is there a user's personal account in some web-based online accounting system and what is available to the user in it?
24. What would you change in the existing chain of HPC center work processes?

Good news — the feedback was received promptly. This illustrates that the questions are vital, and the respondents are willing to improve the discussed situation as it has been done lately [5]. Respondents answered with varying degrees of detail, but interestingly, almost all agreed to do so on condition of anonymity. In other words, the respondents are ready and want to improve something in the current state of things, but would not like it to be visible from the outside. In total, 12 Russian and 2 European centers took part in the survey. All Russian centers are included in the Top50, and foreign centers are in the top of the Top500. They are:

1. Lomonosov Moscow State University HPC center, Moscow, Russia [7, 8]
   `https://parallel.ru/cluster`
2. Joint Supercomputer Center of the Russian Academy of Sciences, Moscow, Russia [9]
   `http://www.jscc.ru/`
3. HSE University, Moscow, Russia [10]
   `https://hpc.hse.ru/hardware/hpc-cluster`
4. Meshcheryakov Laboratory of Information Technologies, Joint Institute for Nuclear Research, Dubna, Russia
   `http://lit.jinr.ru/`
5. SUSU HPC center, Chelyabinsk, Russia
   `https://supercomputer.susu.ru/`
6. Skolkovo Institute of Science and Technology, Moscow, Russia
   `https://www.skoltech.ru/`
7. Siberian SuperComputer Center of ICMMG SB RAS, Novosibirsk, Russia
   `http://www.sscc.icmmg.nsc.ru/`
8. Supercomputer center of UB RAS, Ekaterinburg, Russia
   `http://parallel.uran.ru/`
9. "Polytechnic" Supercomputer center, Saint Petersburg, Russia
   `http://scc.spbstu.ru`
10. UNN Supercomputer center, Nizhniy Novgorod, Russia
    `http://hpc-education.unn.ru/en/`

2023, **24** (1), 1–9. doi 10.26089/NumMet.v24r101

11. Lomonosov Moscow State University CMC Faculty facilities, Moscow, Russia
    http://hpc.cmc.msu.ru/
12. Lomonosov Moscow State University, Big Data Storage and Analysis Center
13. Jülich Supercomputer Centre (JSC), Jülich, Germany
    https://www.fz-juelich.de/en
14. The High Performance Computing Center (HLRS), Stuttgart, Germany
    https://www.hlrs.de/

It can be seen that the scale of both the HPC centers themselves and their systems is presented in all its diversity — from dozens of users to thousands, from tens of TFlops to multipetaflop systems. Let us consider some of the results of the surveys and discussions:

- *Scales of the shared HPC facilities in terms of users.* 2 of 14 centers manage over 2000 users, 1 of 14 — dozens of users, the rest 11 of 14 have hundreds of users
- *"One window" service and personal accounts.* Only 3 of 14 have full-scale implementations, 2 of 14 – under construction, the rest 9 of 14 do not have it at present practice.
- *Resource delegation.* HPC resources are provided to users/logins (4 of 14, the smallest centers), the majority (10 of 14) use the project-oriented workflow.
- *Helpdesk.* There are still centers (3 of 14) which until now do not have a helpdesk and use emails for requests and feedback.
- *Automation and specialized software.* The same 3 of 14 centers do not have specialized software automation for workflow organization and process everything manually or use general script-based automation.
- *Need for changes.* 3 of 14 centers would like to significantly reorganized the workflow organization and means for managing it.

The principles of resource allocation

All contemporary supercomputer centers have long been practicing a resource allocation to a project [6]. This greatly simplifies both the justification of the need for resources when they are requested, and the control over their targeted spending. The idea is depicted in Figures 1 and 2.

Large centers need to conduct an expertise when a project requests resources, and, interestingly, not only when they are provided free of charge. At the European centers, a similar procedure is also required for commercial projects, thereby protecting the reputation of the HPC center.

A periodic, most often annual, procedure for monitoring the use of allocated resources is not implemented everywhere, but in large centers, some form of this monitorning is still is carried out.

General questions

One of the hottest topics is the burden on system administrators. The removal of excess load is given priority by almost every Russian supercomputer center, which indicates their overload with related work, or a shortage of personnel.

Effective provision is not possible without helping users to utilize the available resources by a proper way. Working with user applications and their optimization in some cases is the task of special units, however, not always free of charge, and not in Russian centers.

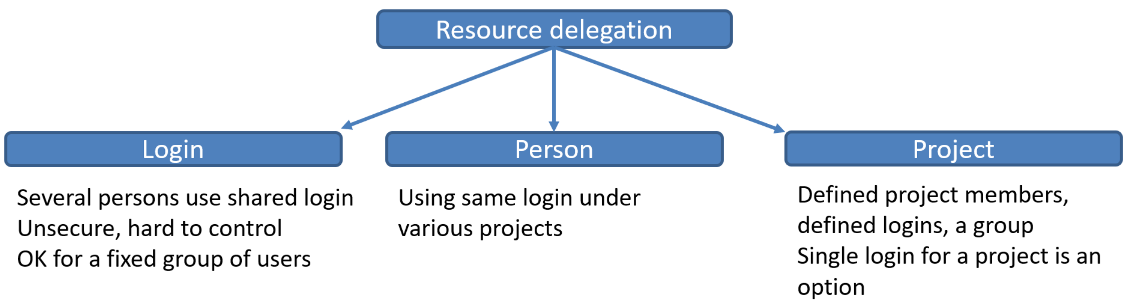

Figure 1. The basic strategies for resource delegation

2023, **24** (1), 1–9. doi 10.26089/NumMet.v24r101

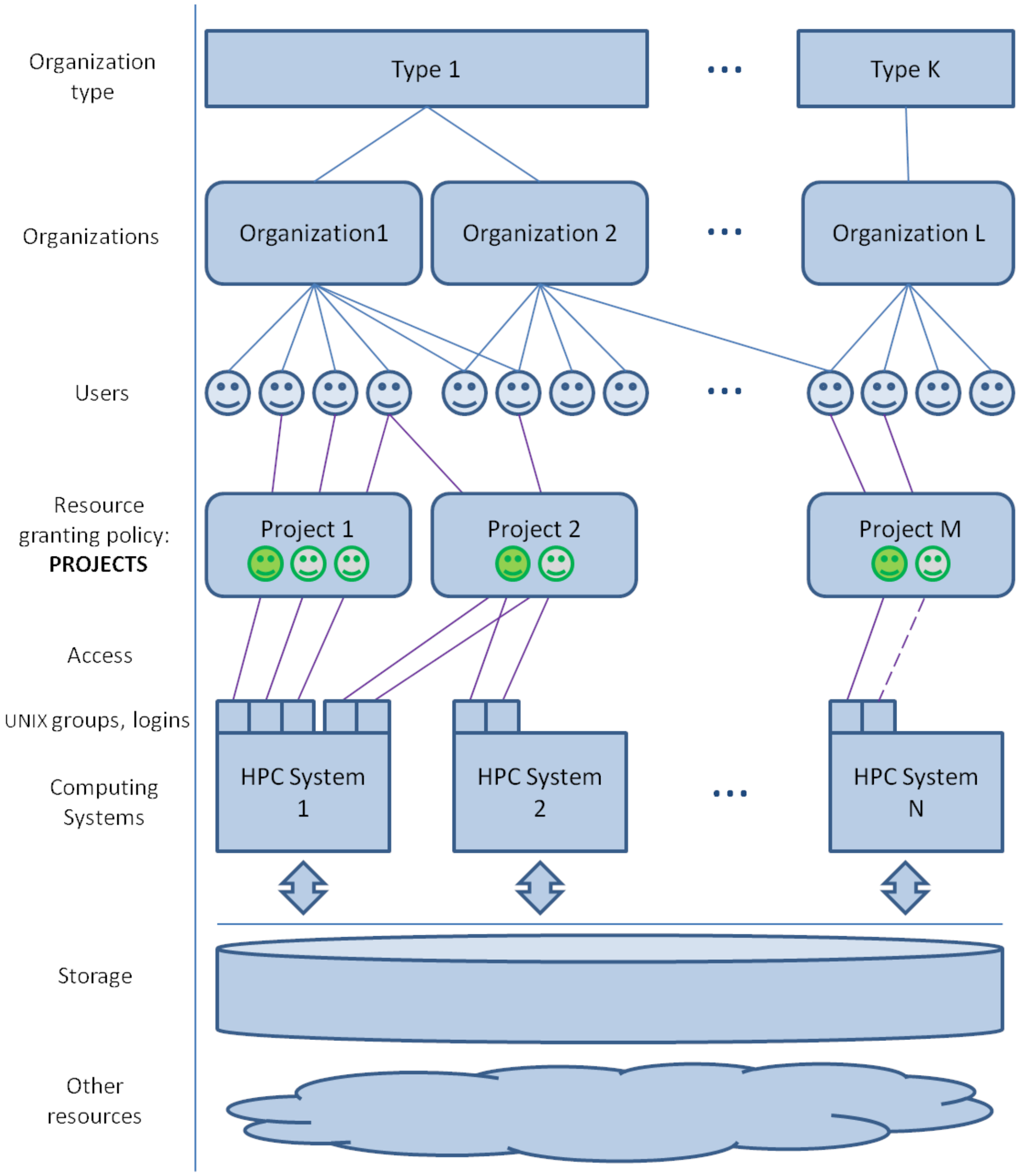

Figure 2. Project-oriented workflow organization scheme [11]

Complete and up-to-date documentation of all software and hardware tools available to users is a necessary foundation for users to start working. This, among other things, reduces the number of unnecessary questions in support.

Direct user support in large centers should be implemented as a flexible tool with rubrics, the ability to involve qualified specialists in the course of resolving issues. Ideally, dedicated experts on the supercomputer center staff should be assigned to communicate with users.

Automation of work processes at all possible stages reduces the impact of the human factor and avoids downtime while waiting for the transition of processes from one state to another. For example, an application for the allocation of resources should automatically continue on its way after a number of conditions are met, for example, keys are entered for access by all project participants.

A "one window service" users should be standard practice for large HPC facility. Such a solution significantly enhances the reliability of interaction with users and in many ways contributes to improving the culture of work in general.

2023, **24** (1), 1–9. doi 10.26089/NumMet.v24r101

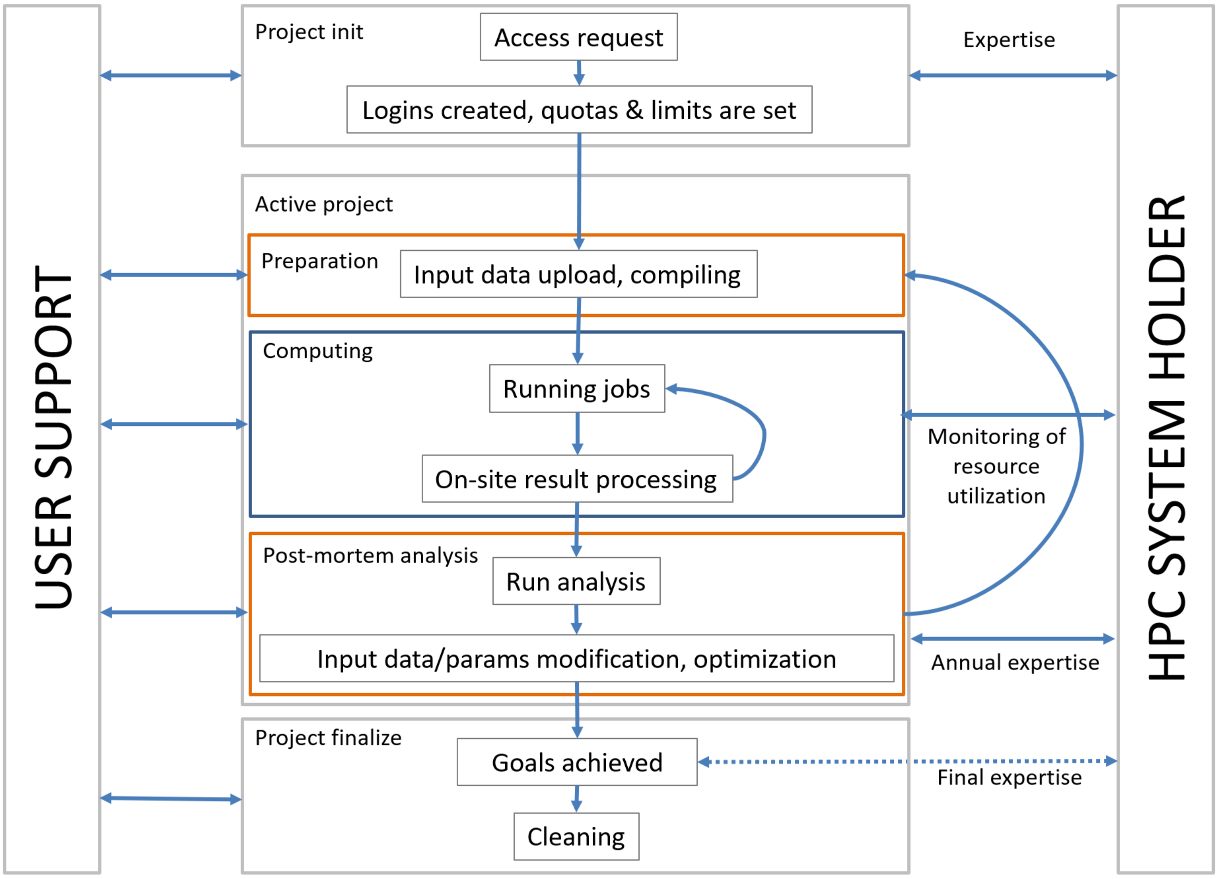

Figure 3. General workflow scheme of HPC centers.

3. Proposed workflow scheme. Based on the analysis of the organization of work in various supercomputer centers, both commercial and public, three main parallel directions can be distinguished that are directly related to obtaining results using computing resources (Figure 3).

The main "user activity" branch.

It is directly related to the work of users, and can be divided into three large phases:

1. The project initial phase, which includes the request for resources and its approval, the establishment of accounts, the activation of licenses and all related preparatory activities.
2. The phase of direct work on the system, in which it is possible to distinguish:
   - The preparatory stage, which includes preparing the input data, compiling the code, setting up the environment.
   - Directly running of computational jobs and preliminary processing of the obtained results without downloading them outside HPC facility.
   - The stage of analysis of the result, which includes the examination of the results, their accuracy, completeness, the necessary adjustment for new runs, including the optimization of the user application and app run conditions, the export of the results.
3. The final phase involves the correct completion of the use of computing resources — clearing the allocated storage for the project, restricting access to participant accounts, and submitting a completion report.

In parallel with the mentioned three phases of user activity, throughout the project, support is being provided by the HPC center, which provides resources in two complementary planes.

Technical support.

Firstly, this is technical support, which should be available to users from the very beginning until the

2023, **24** (1), 1–9. doi 10.26089/NumMet.v24r101

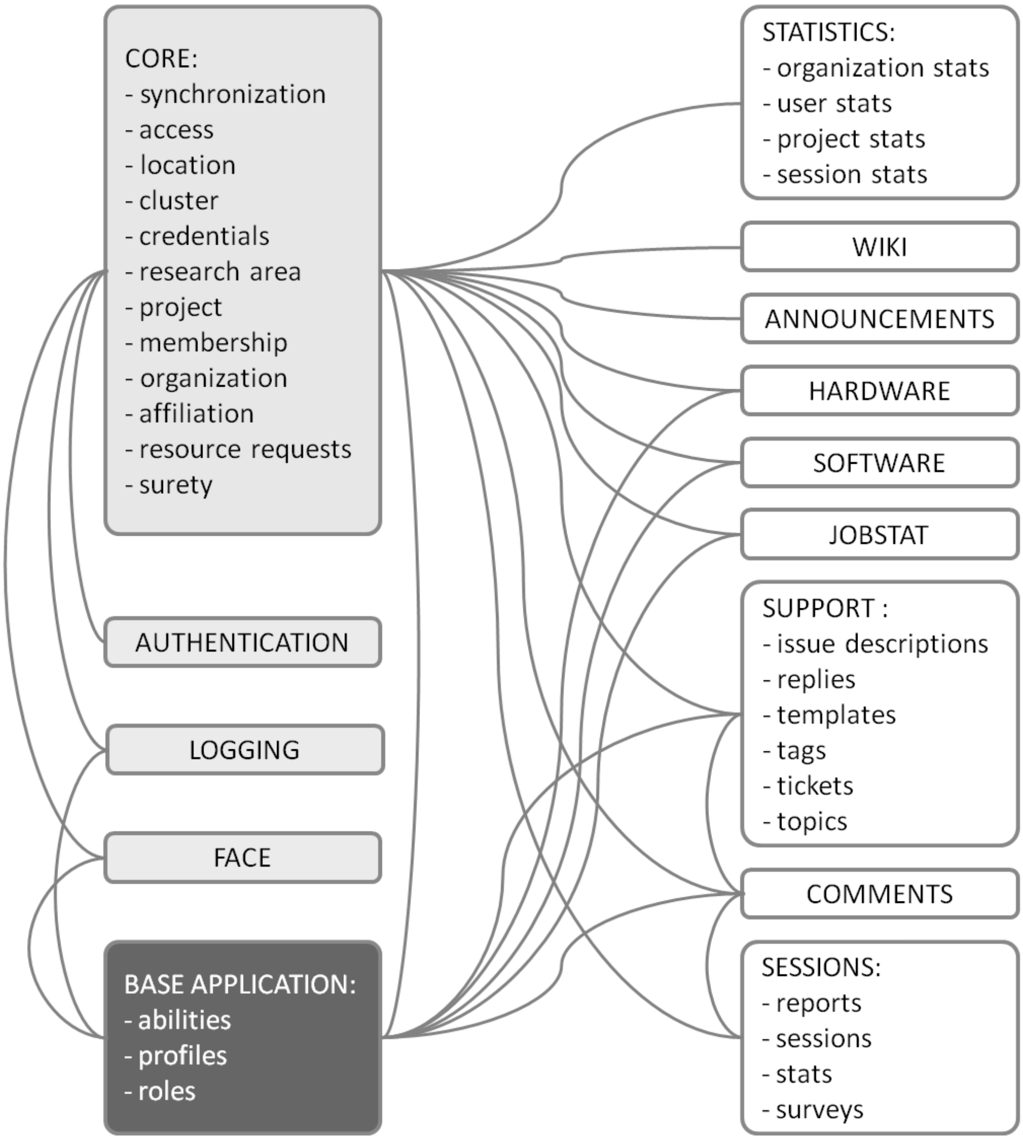

Figure 4. Modular structure of the Octoshell system [11].

end of the project, helping to solve any technical problems related to both the logic of project management and technical problems, for example, loading user data, problems with connecting libraries etc. Note that application optimization is deliberately not included in technical support. This is a separate area of work, and system administrators can be involved in them, but at the same time, a detailed study requires a deep dive into the specifics of the application, the complexity of which removes this task from the list of those that could be solved within the framework of technical support.

Interaction with the system holder.

Secondly, it is interaction with the holder of the system. It is of an organizational and regulatory nature, therefore it is placed separately from technical support.

At the initial stage, this is an examination of the request for resources, adjustment of quotas, etc., in accordance with the specifics of the project and the conditions for the provision of resources.

Further, for an established project, this is a constant monitoring of the project activity, the nature of the use of resources. Here, a mechanism for automatically informing users about anomalies detected during the project can and should be implemented - a change in the nature of resource use, noticed problems with the efficiency of applications, etc. Such a system operates, for example, at the MSU HPC center [12, 13].

Here, for all active projects, regular monitoring of the targeted use of allocated resources may be required in the form, for example, of the provision of annual reports.

For supercomputer centers of different scales, the importance of successful implementation of interaction at each stage is certainly different, but for large HPC facilities with a large number of users and projects, the cost of unsuccessful implementations or omissions can be too high. As a result, this leads to a decrease in the return on the center as a whole and an excessive burden on the administrators and staff of the supercomputer center.

4. Software implementation. The software implementation of the considered model is the Octoshell system developed at the Research Computing Center of Moscow State University, and it continues to be used in the daily practice of the MSU HPC center. To support the described workflows, the system has a modular implementation, and the necessary minimum functionality is realized in the system core (Figure 4).

A fundamentally important feature of the system is the lack of deep integration with system software. The system is, as it were, an add-on over standard tools, which significantly increases reliability. The possibilities of the system and its architecture are considered in detail in [11].

5. Conclusion. The paper presents the main conclusions based on the results of discussions on the organization of work with representatives of the leading HPC centers.

According to the surveys and discussions, the outlined features of organizing the HPC workflow should include at least the following upgrades:

- Reducing the burden on the system administrators.
- Providing means for optimization of user applications.
- Availability of the complete and up-to-date documentation on all software and hardware tools.
- Implementation of effective helpdesk.
- Automation of work processes at all possible stages.
- Providing a “one window service” for users.

The most typical scheme for providing computing resources, which is practiced in most supercomputer centers, is considered. This scheme is fully consistent with the Octoshell HPC operation support system.

References

1. Vad. V. Voevodin, D. I. Shaikhislamov, and D. A. Nikitenko, “How to Assess the Quality of Supercomputer Resource Usage,” Supercomput. Front. Innov. **9** (3), 4–18 (2022). doi 10.14529/jsfi220301.
2. R. McLay, K. W. Schulz, W. L. Barth, and T. Minyard, “Best Practices for the Deployment and Management of Production HPC Clusters,” in *Proc. Int. Conf. for High Performance Computing, Networking, Storage and Analysis, Seattle, USA, November 12–18, 2011* (ACM Press, New York, 2011), doi 10.1145/2063348.2063360.
3. S. Varrette, P. Bouvry, H. Cartiaux, and F. Georgatos, “Management of an Academic HPC Cluster: The UL Experience,” in *Proc. Int. Conf. on High Performance Computing & Simulation, Bologna, Italy, July 21–25, 2014* (IEEE Press, New York, 2014), pp. 959–967. doi 10.1109/HPCSim.2014.6903792.
4. S. Varrette, E. Kieffer, and F. Pinel, “Optimizing the Resource and Job Management System of an Academic HPC & Research Computing Facility,” in *Proc. 21st Int. Symposium on Parallel and Distributed Computing, Basel, Switzerland, July 11–13, 2022* (IEEE Press, New York, 2022), pp. 129–137. doi 10.1109/ISPDC55340.2022.00027.
5. Vad. V. Voevodin, R. A. Chulkevich, P. S. Kostenetskiy, et al., “Administration, Monitoring and Analysis of Supercomputers in Russia: a Survey of 10 HPC Centers,” Supercomput. Front. Innov. **8** (3), 82–103 (2021). doi 10.14529/jsfi210305.
6. A. V. Paokin and D. A. Nikitenko, “Unified Approach for Provision of Supercomputer Center Resources,” Vestn. Yuzhn. Ural. Gos. Univ. Ser. Vychisl. Mat. Inf. **11** (1), 5–14 (2022). doi 10.14529/cmse220101.
7. V. Voevodin, A. Antonov, D. Nikitenko, et al., “Lomonosov-2: Petascale Supercomputing at Lomonosov Moscow State University,” in *Contemporary High Performance Computing* (CRC Press, Boca Raton, 2019), Vol. 3, pp. 305–330. doi 10.1201/9781351036863-12.
8. Vl. V. Voevodin, A. S. Antonov, D. A. Nikitenko, et al., “Supercomputer Lomonosov-2: Large Scale, Deep Monitoring and Fine Analytics for the User Community,” Supercomput. Front. Innov. **6** (2), 4–11 (2019). doi 10.14529/jsfi190201.

2023, **24** (1), 1–9. doi 10.26089/NumMet.v24r101

9. G. I. Savin, B. M. Shabanov, P. N. Telegin, and A. V. Baranov, "Joint Supercomputer Center of the Russian Academy of Sciences: Present and Future," Lobachevskii J. Math. **40** (11), 1853–1862 (2019). doi 10.1134/S1995080219110271.
10. P. S. Kostenetskiy, R. A. Chulkevich, and V. I. Kozyrev, "HPC Resources of the Higher School of Economics," J. Phys.: Conf. Ser. **1740** (1), Article 012050 (2021). doi 10.1088/1742-6596/1740/1/012050.
11. D. A. Nikitenko, Vad. V. Voevodin, and S. A. Zhumatiy, "Driving a Petascale HPC Center with Octoshell Management System," Lobachevskii J. Math. **40** (11), 1817–1830 (2019). doi 10.1134/S1995080219110192.
12. P. Shvets, Vad. Voevodin, and D. Nikitenko, "Approach to Workload Analysis of Large HPC Centers," in *Communications in Computer and Information Science* (Springer, Cham, 2020), Vol. 1263, pp. 16–30. doi 10.1007/978-3-030-55326-5_2.
13. D. A. Nikitenko, P. A. Shvets, and V. V. Voevodin, "Why do Users Need to Take Care of Their HPC Applications Efficiency?," Lobachevskii J. Math. **41** (8), 1521–1532 (2020). doi 10.1134/s1995080220080132.

Received
December 28, 2022

Accepted for publication
January 9, 2023

Information about the author

Dmitry A. Nikitenko — Ph.D., Leading Scientist; Lomonosov Moscow State University, Research Computing Center, Leninskie Gory, 1, building 4, 119234, Moscow, Russia.