

The Main Scientific and Technical Problems of Using Hybrid HPC Clusters in Materials Science¹

K. I. Volovich^{a,*} and S. A. Denisov^{a,**}

^aFederal Research Center “Computer Science and Control” of the Russian Academy of Sciences, Moscow, 119333, Russia

*e-mail: KVolovich@frcsc.ru

**e-mail: SDenisov@frcsc.ru

Received December 26, 2019; revised February 19, 2020; accepted March 3, 2020

Abstract—The article discusses the use of hybrid HPC clusters for the execution of software designed to calculate the electronic structure and atomic scale materials modeling. Modern software systems, which are designed to solve the problems of materials science, use the capabilities of various hardware computing accelerators to increase productivity. The use of such computing technologies requires the adaptation of application program code to hybrid computing architectures, which include classic central processing units (CPUs) and specialized graphics accelerators (GPUs). The use of large computing hybrid systems requires the development of methods for ensuring the workloading of such computing systems that will allow efficient use of computing resources and avoid equipment downtime. First of all, these methods should allow parallel execution of user applications using computational accelerators. However, in practice, software environments designed to solve application problems cannot be deployed in the same computing environment due to software incompatibility. In order to overcome this limitation and ensure the parallel execution of diverse types of materials science tasks, the creation of individual task execution environments based on virtualization technologies and cloud technologies. The continuation of virtualization technologies and the provision of cloud services is the construction of digital platforms. The article proposes the use of a digital platform for hosting scientific materials science services that provide calculations using various application software systems. Digital platforms make it possible to provide a unified user interface to scientific materials science services. The platform provides opportunities for finding the necessary scientific services, transferring source data and results between users, the platform and hybrid high-performance clusters.

Keywords: high-performance computing cluster, hybrid architecture, graphics accelerator, electronic structure calculations, quantum-mechanical molecular dynamics, VASP, Quantum ESPRESSO

DOI: 10.1134/S1063739720080090

INTRODUCTION

The current trend in science is the increasing role of computer calculations [1]. Modern software systems designed to solve problems of materials science and other fields of science, require huge computational resources [2–5]. They are most effective when deployed in HPC clusters, which have high performance and allow you to solve problems in a reasonable time with sufficient accuracy [6, 19].

In HPC clusters, there is an active introduction of hybrid architectures. Special components are added to computer systems—GPU, which can significantly improve the performance of calculations and reduce the time of scientific calculations.

Materials science software systems use graphics accelerators for calculations. As an example, the

software systems VASP (The Vienna Ab initio Simulation Package) and Quantum ESPRESSO. These software systems exhibit the greatest efficiency in the exclusive use of computing resources: RAM, CPU, GPU.

Note that the operation of a hybrid HPC cluster poses the problem of managing resources and dividing them among a group of users. Technologies need to be developed that provide resources for Materials Science applications for different users and scientific teams.

A modern approach to workload management is the use of virtualization and cloud technologies. Cloud technologies allow us to provide users with SaaS and PaaS services. It is advisable to provide science teams with applied materials science systems as cloud services.

The aim of the work is to consider the main problems and approaches to the provision of material sci-

¹This article was prepared based on a report presented at the 1st International Conference on “Mathematical Modeling in Materials Science of Electronic Components” (Moscow, 2019).

ence software in the environment of a hybrid high-performance cluster using cloud technologies.

ASPECTS OF THE USE OF CLOUD TECHNOLOGIES IN THE HYBRID HPC CLUSTER IN THE PERFORMANCE OF MATERIALS SCIENCE TASKS

The use of cloud technologies is today a current trend in all areas related to information technology [7, 12]. The provision of hybrid high-performance computing services in the field of materials science is no exception.

Currently, there are a number of software systems that perform problem solving in various areas of materials science. Such systems as ABINIT, VASP, Quantum ESPRESSO, LAMMPS, MEEP, MPB, OpenFOAM can run on hybrid HPC clusters, however, it requires the preparation of a runtime environment, the workload management system setup, the organization of parallel computing [8].

The computing environment requires greater flexibility and adaptability to changing requirements from applications and user tasks.

At the same time, it is possible to identify the main problems that arise when using cloud technologies:

- the problem of deploying the code of the software science materials system in a hybrid HPC cluster environment;
- the problem of adapting the user code to the hybrid computer;
- the problem of creating an individual environment for the execution of tasks [9].

To solve these problems, it is necessary to develop methods and algorithms for the adaptation and execution of applications for users on a hybrid computing cluster that uses specialized computing accelerators to perform calculations. Solutions for adapting software code in languages of various levels to the target platform of a hybrid computing cluster, algorithms for providing computing resources to tasks in a multitasking environment, as well as methods for creating an adapted task execution environment using virtualization technologies require scientific study and are relevant at the current stage of theory and development. practices of providing computing resources of a hybrid architecture to research teams.

The use of cloud technologies, on the one hand, provides a number of advantages associated with a unified approach to providing services to users, and, on the other hand, gives rise to problems caused by the need to provide a wide range of services and adapt the computing environment of the calculator to the needs of each user task.

The issues of providing resources for such systems and the simultaneous execution of tasks that are different in scientific orientation require solving from the

point of view of developing efficient algorithms for organizing the computational process. It is necessary to ensure efficient work loading of the hybrid cluster with tasks of various types, which allow minimizing equipment downtime and reduce the time for performing scientific tasks.

A separate scientific study requires the adaptation of application software code to the architecture of the hybrid computing system. This task should include the classification of software according to the degree of mobility, approaches to translating software code into technologies that allow the use of computing accelerators and the creation of an individual environment for performing a scientific task on the resources of a hybrid computing cluster.

As a result of solving problems of creating an individual execution environment and deploying system software, the materials science software package and user applications, it can create services like SaaS and PaaS. Figure 1 shows an example of the use of scientific teams to the materials science services. For each request to the service, an individual runtime environment is created in which the software system for scientific calculations is deployed. Virtualization technologies provide independent parallel operation of calculated jobs in a hybrid HPC cluster.

As a basis for the deployment of virtual environments, images of virtual machines containing the complete set of software necessary for solving an application problem can be used. In this case, the user is provided with a service like SaaS.

Another possibility to provide a virtual environment is to deploy a framework in it using which the user deploys his own software designed to solve materials science problems. In this case, the user receives the PaaS service and, on its basis, creates a computing environment for solving an applied problem. Note that in the future, based on such decisions, SaaS services can be deployed.

IMPROVING THE PERFORMANCE OF MATERIALS SCIENCE TASKS BASED ON GRAPHICS ACCELERATORS

The use of graphic accelerators significantly improves the performance of solving scientific problems and reduces the execution time of applications. It should be noted that grade enhancement performance depends on the type of accelerator and the type of application [10]. However, in the general case, the use of hybrid HPC clusters leads to an increase in the quality of computing services.

The work [11] addresses the issues of providing high-performance resources as a cloud service, analyzes open cloud platforms (Amazon, IBM, Google), provides approaches to the use of specialized computing accelerators to improve the performance of computing systems to expand cloud platform services,

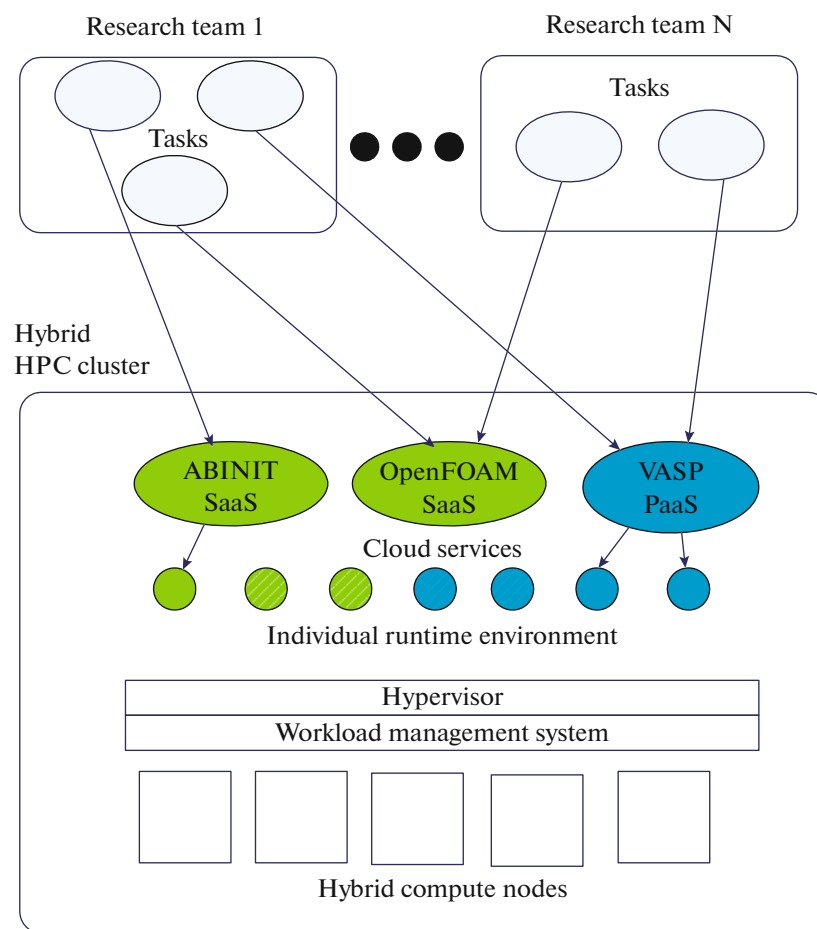


Fig. 1. Cloud services for materials science in hybrid HPC.

which requires the use of hybrid architectures building computing systems. With regard to the problems of materials science, it can be noted that most software systems designed for carrying out calculations use GPU resources to speed up calculations.

Also, there is a tendency to adapt algorithms for solving applied problems to the conditions of execution on computing facilities of a hybrid architecture equipped with accelerators of calculations.

Computational algorithms differ depending on the application, but for all areas the use of mechanisms of hybrid high-performance clusters is typical:

—MPI is a technology that allows you to use simultaneous parallel processes and organize the exchange of data between them;

—CUDA—technology access and perform calculations on graphics accelerators.

Thus, the development projects available for research teams aimed at solving applied problems can be transferred to the resources of a hybrid high-performance computing cluster, however, refinement is required to take into account the use of new technologies.

When transferring, naturally, the expediency factor must be taken into account, showing how effectively the problem will be solved using hybrid technologies and what costs will be required for the processing of codes and algorithms.

However, the tendency to introduce hybrid technologies shows that the advantages provided by hybrid computers are so great that there is a migration of existing software and algorithms to these technologies.

INTEGRATION OF MATERIAL SCIENCE SERVICES THROUGH A DIGITAL PLATFORM

The current trend in providing computing services is cloud computing and a platform-based approach to providing services.

A digital platform for scientific research is an information object that allows service providers to post information on scientific services and provide access to these services through a single information portal of a digital platform. Accordingly, consumers of scientific services can receive information about available services from a single registry of a digital platform and order, execute and receive results through a digital

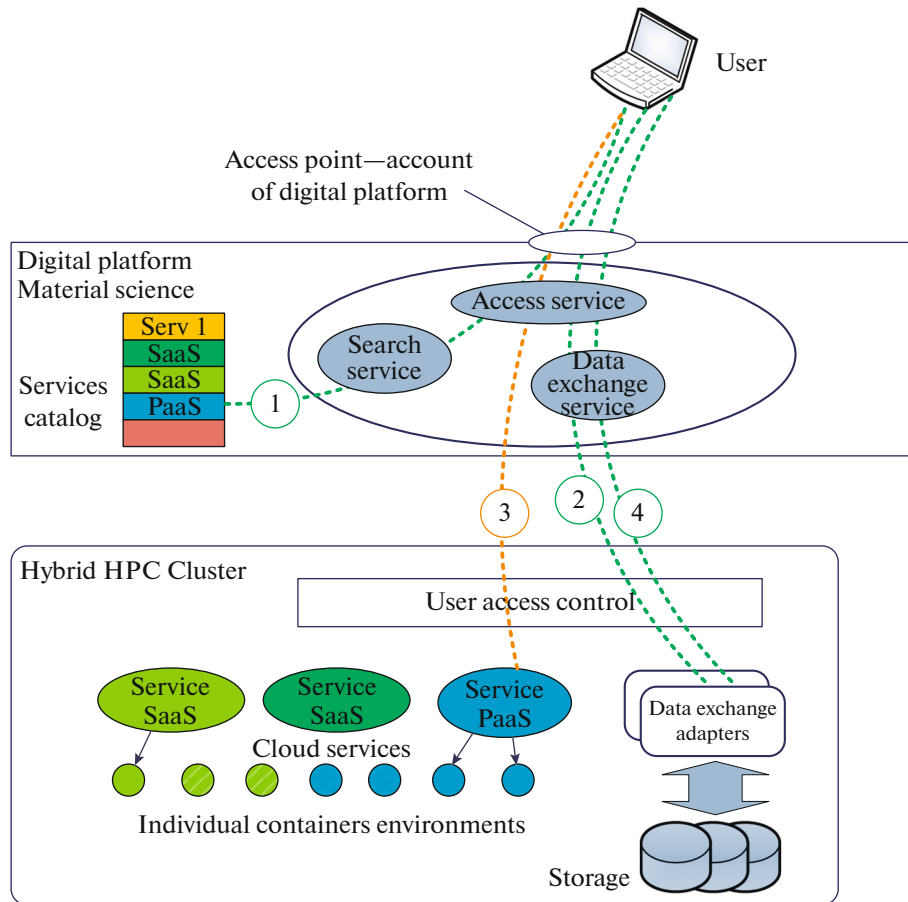


Fig. 2. Access to the HPC cluster through the digital platform.

platform portal. The aspects of digital platforms for scientific research is described in detail in [12, 13].

Thus, the provision of hybrid computing services as a service of a digital platform is a logical continuation of the concept of cloud computing and platform technologies [14, 15].

In the field of material science, this model allows the user a registry of services that provide solutions to problems in this area.

Figure 2 shows an example of user access to cloud computing resources of the digital materials science platform.

The search service provides a search in the registry of ready-made services developed using the technology described above. It is possible to access these services through a single entry point of a digital platform using common SSO authentication.

PaaS is very convenient for research teams to deploy and operate their own software on the platform, as well as frameworks and integrated environments.

Examples of the use of PaaS in the field of material science include the creation and operation of calcula-

tion services of electronic-structure Quantum ESPRESSO [16].

In the field of the provision of scientific services, a striking example of PaaS is the provision of services of hybrid high-performance computing systems for performing calculations that require large computing power. Such calculations are used in many areas of modern science and the services of supercomputer centers are in demand from scientific teams of various fields of knowledge [17]. We note that in the general case, the service of a supercomputer center for the provision of calculators cannot be explicitly considered as PaaS in terms of cloud computing. Algorithms for resource allocation, allocation of computers and user services for supercomputers differ from the classical work of cloud providers. In particular, to accelerate the processes of computing in supercomputers, virtualization and parallel operation of one computing unit are not used in the interests of a group of users. It uses batch processing of jobs, job queues and other methods of organizing the computational process, which allow optimizing the operation of the supercomputer [18].

The publication and provision of services of hybrid computing systems in a digital platform is carried out

through the implementation of information interaction protocols provided for in the digital platform. To do this, specialized interface modules, adapters, are developed that allow the exchange of data between digital platform and computer system management services. It should be borne in mind that the management service of the digital platform does not perform the functions of controlling the computing process, but provides tasks for the calculations and the initial data received by the digital platform.

Also, through the means of a digital platform, on-line user access to the tools of the computer system can be carried out to prepare and test calculation tasks. Thus, a user of a hybrid computing cluster receives cloud computing services (SaaS, PaaS) through a unified interface of a digital platform for scientific research.

CONCLUSIONS

Materials science software systems require large amounts of computational resources. It is advisable to perform them in hybrid HPC cluster environments. The modern effective solution for the implementation of these software systems is the creation of individual execution environments based on virtualization technologies. Container technologies allow you to create isolated virtual objects containing all the necessary software to perform materials science calculations. The user may be provided with a fully prepared environment for performing calculations. In this case, he receives the SaaS cloud service, has the ability to download data and make calculations. Another option for using cloud computing services is PaaS. In this case, the user receives a software environment containing the necessary software tools. He deploys his own software designed for calculating materials science, configures licenses, software environment and performs calculations.

The preferred mode of providing resources of the hybrid HPC cluster for solving problems of materials science is PaaS based on container technology. In this case, the cluster resources can be used most fully and flexibly, as well as the capabilities of the materials science software for working with graphics accelerators.

ACKNOWLEDGMENTS

The experiments on the deployment of individual runtime environments for software packages of materials science were carried out using computing resources of shared research facilities CKP "Informatics" of FRC CSC RAS [20].

FUNDING

The research is partially supported by the Russian Foundation for Basic Research (projects 18-29-03100, 19-29-03051).

REFERENCES

1. Abramov, S.M. and Lilitko, E.P., Current state and development prospects of high-end HPC system, *Inform. Technol. Comput. Syst.*, 2013, no. 2, pp. 6–22.
2. Zhuravlev, A.A., Reviznikov, D.L., and Abgaryan, K.K., The method of discrete elements with an atomic structure, in *Proceedings of the 21st International Conference on Computational Mechanics and Modern Applied Software Systems (VMSPSPS'2019), Moscow, 2019*, pp. 59–61.
3. Mikurova, V.S., Starlings. Creating a generalized model for predicting the inhibition of influenza virus neuraminidase of various strains, *Biomed. Chem.*, 2018, vol. 64, no. 3, pp. 247–252.
4. Mikurova, V.S. and Skvortsov, O.A., Raevsky computer assessment of the selectivity of inhibition of muscarinic receptors M1-M4, *Biomed. Chem.: Res. Methods*, 2018, vol. 1, no. 3.
5. Vouzis, P.D. and Sahinidis, N.V., GPU-BLAST: using graphics processors to accelerate protein sequence alignment, *Bioinformatics*, 2011, vol. 27, no. 2, pp. 182–188.
6. Gorchakov, A.Yu. and Malkova, V.U., Comparison of Intel Core-17, Intel Xeon, Intel Xeon Phi and IBM power 8 processors using the example of initial data recovery, *Int. J. Open Inform. Technol.*, 2018, vol. 6, no. 4.
7. Volkov, S. and Sukhoroslov, O., Simplifying the use of clouds for scientific computing with Everest, *Proc. Comput. Sci.*, 2017, vol. 119, pp. 112–120.
8. Gorchakov, A.Yew., Using openmp to implement the multithreaded method of uneven coatings, in *Advanced Information Technologies (PIT 2018), Proceedings of the International Scientific and Technical Conference, 2018*, pp. 613–617.
9. Volovich, K.I., Denisov, S.A. and Malkovsky, S.I., The formation of an individual modeling environment in a hybrid high-performance computing complex, in *Proceedings of the 1st International Conference on Mathematical Modeling in Materials Science of Electronic Components MMMEC-2019*, Moscow: MAX Press, 2019, pp. 21–24.
10. Afanasyev, I. and Voevodin, V., The comparison of large-scale graph processing algorithms implementation methods for Intel KNL and NVIDIA GPU, *Commun. Comput. Inform. Sci.*, 2017, vol. 793, pp. 80–94.
11. Ding, F., Mey, D., Wienke, S., Zhang, R., and Li, L., A study on today's cloud environments for HPC applications, in *Proceedings of the 3rd International Conference on Cloud Computing and Services Science CLOSER 2013, Aachen, Germany, May 8–10, 2013*, Berlin: Springer, 2014, pp. 114–127.
12. Volovich, K.I., Zatsarinnyy, A.A., Kondrashev, V.A., and Shabanov, A.P., Scientific research as a cloud service, *Sist. Sredstva Inform.*, 2017, vol. 27, no. 1, pp. 73–84.
13. Zatsarinny, A.A., Gorshenin, A.K., Kondrashev, V.A., Volovich, K.I., and Denisov, S.A., Toward high performance solutions as services of research digital platform, in *Proceedings of the 13th International Symposium on Intelligent Systems INTELS'18, St. Petersburg, Russia, October 22–24, 2018, Elsevier Proc. Comput. Sci.*, 2019, no. 150, pp. 622–627.
14. Zatsarinny, A.A., Gorshenin, A.K., Volovich, K.I., Kolin, K.K., Kondrashev, V.A., and Stepanov, P.V.,

- Management of scientific services as the basis of the national digital platform “Science and Education,” *Strateg. Priorities*, 2017, no. 2 (14), pp. 103–113.
15. Kondrashev, V.A. and Volovich, K.I., Service management of a digital platform on the example of high-performance computing services, in *Proceedings of the International Scientific Conference, Voronezh, September 3–6, 2018*.
 16. Kartsev, A., Malkovsky, S.I., Volovich, K.I., and Sorokin, A.A., The study of the performance and scalability of the Quantum ESPRESSO package in the study of low-dimensional systems on hybrid computing systems, in *Proceedings of the 1st International Conference on Mathematical Modeling in Materials Science of Electronic Components MMMEC-2019*, Moscow: MAX Press, 2019, pp. 18–21.
 17. Berriman, G.B., Deelman, E., Juve, G., Rynge, M., and Vockler, J.-S., The application of cloud computing to scientific workflows: A study of cost and performance, *Phil. Trans. R. Soc. London, Ser. A*, 2013, vol. 371, p. 20120066. Iss. 1983.
<https://doi.org/10.1098/rsta.2012.0066>
 18. Yakobovsky, M.V., Bondarenko, A.A., Vyrodov, A.V., Grigoryev, S.K., Kornilina, M.A., Plotnikov, A.I., Polyakov, S.V., Popov, I.V., Bubbles, D.V., and Sukov, S.A., Cloud service for solving multiscale nanotechnology problems on clusters and supercomputers, *News SFU, Tech. Sci.*, 2016, no. 12 (185).
 19. Gorchakov, A.Yu. and Posypkin, M.A., Comparison of multi-threaded implementations of the branch and bound method for multicore systems, *Mod. Inform. Technol. IT Educ.*, 2018, vol. 14, no. 1, pp. 138–148.
 20. Regulations of CKP ‘Informatics.’ <http://www.frccsc.ru/ckp>. Accessed January 22, 2020.