

Облачные сервисы для научных высокопроизводительных вычислений на базе платформы Proxmoх*

А. В. БАРАНОВ[†], Е. А. КИСЕЛЁВ

Межведомственный суперкомпьютерный центр РАН — филиал ФНЦ НИИСИ РАН,
Москва, Россия

[†]Контактный e-mail: antbar@mail.ru

Организация облачных сервисов для высокопроизводительных вычислений затруднена, во-первых, по причине высоких накладных расходов на виртуализацию, во-вторых, из-за специфики систем управления заданиями и ресурсами в научных суперкомпьютерных центрах. В настоящей работе рассмотрен подход к построению облачных сервисов видов PaaS и SaaS, основанных на совместном функционировании облачной платформы Proxmoх VE и системы управления прохождением параллельных заданий, применяемой в качестве менеджера ресурсов в Межведомственном суперкомпьютерном центре РАН.

Ключевые слова: суперкомпьютер, облачные сервисы, Proxmoх, накладные расходы, СУППЗ.

Библиографическая ссылка: Баранов А.В., Киселёв Е.А. Облачные сервисы для научных высокопроизводительных вычислений на базе платформы Proxmoх // Вычислительные технологии. 2019. Т. 24, № 6. С. 5–12.
DOI: 10.25743/ICT.2019.24.6.002.

Введение

Термин “облачные сервисы” ежегодно упоминается в отчетах крупных отечественных и зарубежных ИТ-компаний. В 2018 г. было отмечено, что рынок облачных сервисов растет в 4.5 раза быстрее, чем вся ИТ-отрасль [1]. При этом сегмент программного обеспечения, предоставляемого в качестве услуги (SaaS), остается самым крупным на облачном рынке с долей в 62.4% [1]. В России темп роста рынка облачных услуг соответствует мировому уровню и зачастую превышает прогнозы аналитиков [2]. Так, по прогнозам компании SAP СНГ, сделанным в 2017 г., предполагалось что в 2018 г. рынок облачных услуг России составит 34.41 млрд руб., однако согласно оценке iKS-Consulting по итогам 2018 г. российский рынок облачных услуг уже достиг 68.4 млрд руб. и прогнозируется его дальнейшее увеличение (рис. 1).

Иначе обстоит дело в отечественных научных вычислительных центрах коллективного пользования, прежде всего — суперкомпьютерных. Здесь пользователям по-прежнему предоставляется консервативный перечень услуг для организации вычислений, и причина не в отсталости самих центров, а в уникальности и сложности супер-

*Title translation and abstract in English can be found on page 12.

© ИВТ СО РАН, 2019.

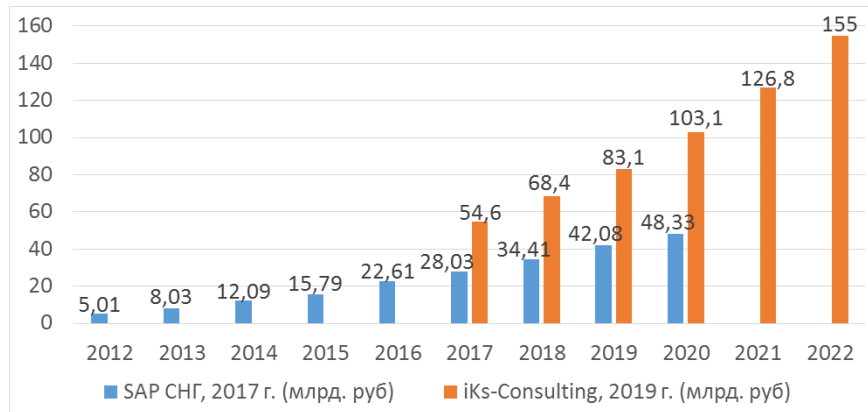


Рис. 1. Темп роста рынка облачных услуг в России

компьютеров и специфике организации вычислительного процесса. Современный суперкомпьютер — это, как правило, кластер, состоящий из объединенных высокоскоростной сетью вычислительных узлов. Для возможности осуществления расчетов и получения доступа к узлам суперкомпьютера пользователь должен оформить так называемое задание (параллельное задание, вычислительное задание) — информационный объект, включающий расчетную программу, требования к ресурсам (сколько, каких узлов и на какое время требуется для расчетов) и в ряде случаев исходные данные. Обработку потока различных заданий от разных пользователей осуществляют специальные системы управления заданиями [3], предоставляющие заданиям инфраструктуру с типовым набором программного обеспечения для организации параллельных вычислений (IaaS), в некоторых случаях — необходимый набор средств разработки и выполнения параллельных приложений (PaaS).

Создание облачных сервисов более высокого уровня (PaaS, SaaS) в IT-индустрии связано с применением технологий виртуализации, что, как показано в работах [4, 5], влечет не приемлемые для высокопроизводительных вычислений накладные расходы для большинства параллельных программ. Однако достаточно широкий класс суперкомпьютерных приложений, например разрабатываемых и применяемых в образовательном процессе, не столь критичны к накладным расходам на виртуализацию. Соответствующие задания могут быть представлены не в виде расчетных программ, а в виде виртуальных машин или контейнеров. Для возможности обработки таких заданий совместно с обычными, традиционными заданиями в настоящей работе предлагается подход, основанный на внедрении API-платформы виртуализации Proxmox VE [6] в состав отечественной Системы управления прохождением параллельных заданий (СУППЗ) [7], применяемой в качестве системы управления заданиями (СУЗ) в Межведомственном суперкомпьютерном центре РАН (МСЦ РАН).

1. Структура облачной среды для высокопроизводительных вычислений

Применяемые для построения PaaS- и SaaS-сервисов виртуальные машины и контейнеры позволяют создавать необходимое для каждого пользователя программное окружение, но они не приемлемы для большинства суперкомпьютерных приложений. Помимо

высоких накладных расходов на виртуализацию [4] среди причин отказа называют [5] уникальность суперкомпьютерных ресурсов, которая не всегда поддерживается гипервизорами, и необходимость непосредственного доступа пользовательского приложения к суперкомпьютерным ресурсам для его максимальной оптимизации. Для сохранения преимуществ технологии виртуализации при построении облачных сервисов и минимизации накладных расходов при выполнении параллельных заданий пользователей необходимо сопряжение облачных платформ и СУЗ суперкомпьютера. В работе [5] в качестве примера такого сопряжения предложено внедрение драйвера СУЗ в программный стек облачной платформы OpenStack. Альтернативный подход состоит во внедрении API-платформы виртуализации в состав СУЗ. Известны решения по интеграции СУЗ и контейнеров Linux. Так, в СУЗ SLURM реализован модуль для запуска на выделенных заданию вычислительных ресурсах контейнеров Linux [8]. В работе [9] исследовано влияние контейнерной виртуализации на время выполнения параллельных программ, что позволило [10] интегрировать в СУППЗ поддержку пользовательских заданий, представленных в виде контейнеров.

Рассмотрим аналогичное [10] решение, предлагаемое авторами не только для контейнеров, но и для виртуальных машин. На основе этого решения в МСЦ РАН разработана технология построения облачных сервисов для проведения научных исследований, ключевыми особенностями которой являются:

- использование отечественного программного обеспечения либо программного обеспечения, распространяемого в открытых исходных текстах;
- возможность одновременного функционирования трех типов сервисов (IaaS, PaaS и SaaS) на одной вычислительной системе;
- возможность объединения нескольких суперкомпьютерных систем в единое облако.

На базе предложенной технологии создана облачная среда для проведения научных исследований, структура которой состоит из трех уровней. Каждый уровень представляет свой тип облачного сервиса: IaaS, PaaS и SaaS (рис. 2). Пользователь может одновременно работать как с одним, так и с несколькими типами сервисов.

Доступ пользователей к IaaS-сервисам осуществляется через командный интерфейс СУППЗ по протоколу SSH. Пользователю выделяется необходимое количество вычис-

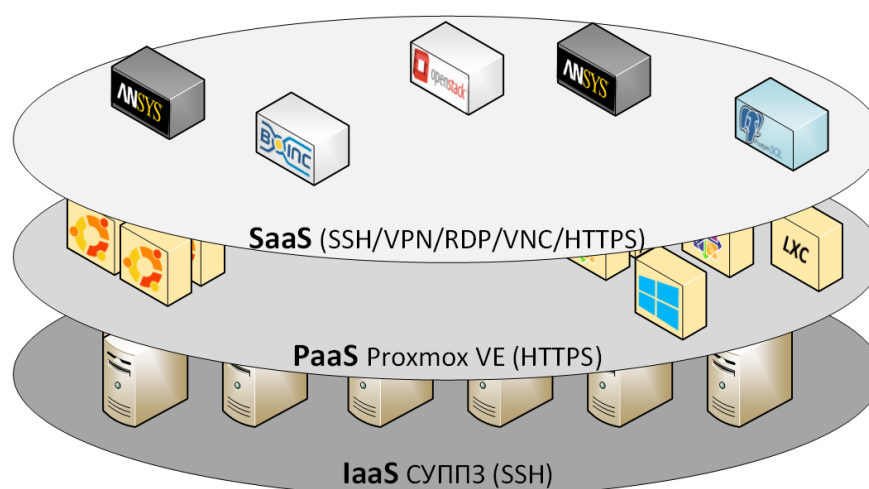


Рис. 2. Структура облачной среды для проведения научных исследований

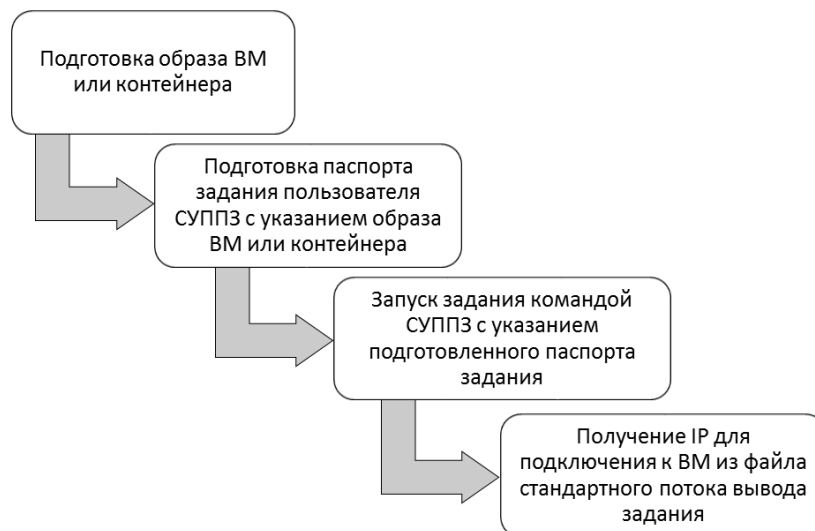


Рис. 3. Схема запуска виртуальных машин и контейнеров через СУППЗ

лительных узлов и предоставляется беспарольный доступ на каждый из них. Время доступа определяется через командный интерфейс СУППЗ и ограничивается заданными администратором настройками. Для каждого пользователя автоматически создается проектный каталог, расположенный на сетевом дисковом пространстве, и предоставляется набор средств разработки MPI-, OpenMP-, Python-программ, отладки и контроля версий. Стандартные потоки ввода, вывода и ошибок перенаправляются в рабочий каталог задания пользователя.

РaaS-сервисы функционируют под управлением платформы виртуализации Proxmox VE [6]. По запросу пользователя СУППЗ выделяет требуемое количество вычислительных узлов и сохраняет их IP-адреса в стандартный поток вывода задания. Выделенные узлы образуют виртуальный вычислительный кластер, на котором пользователь может разворачивать как виртуальные машины KVM, так и контейнеры. Доступ к виртуальному кластеру и управление виртуальными машинами (VM) осуществляются пользователем по протоколу HTTPS через браузер. В качестве URL-адреса для подключения к виртуальному кластеру может быть использован любой IP-адрес из выделенных при старте задания. В автоматическом режиме выделения вычислительных узлов СУППЗ ограничивает время аренды виртуального кластера, но по запросу пользователей администратор может увеличить время аренды путем блокировки узлов суперкомпьютера после старта задания. Схема запуска виртуальных машин представлена на рис. 3.

SaaS-сервисы представляют собой виртуальные машины или контейнеры с запущенными сервисами: удаленной визуализации результатов моделирования и разработки САД-моделей для проведения расчетных экспериментов, системы коллективной разработки программного обеспечения, организации вычислений с распараллеливанием по данным, почтовые службы, базы данных, VPN, виртуальные лаборатории, удаленные рабочие столы и др. Виртуальные машины и контейнеры функционируют в круглосуточном режиме на выделенных администраторами МСЦ РАН вычислительных модулях, объединенных в отказоустойчивый виртуальный кластер средствами Proxmox. В зависимости от вида SaaS-сервиса применяется свой протокол доступа: SSH, RDP, VNC, HTTPS или VPN. Для организации контроля доступа пользователей к каждому SaaS-сервису используется LDAP или Active Directory.

2. Оценка влияния компонентов виртуальной среды Proxmox VE на выполнение MPI-программ

В рамках настоящего исследования авторами проведена оценка влияния компонентов Proxmox VE, установленных на вычислительных узлах суперкомпьютера, и виртуальной среды на время выполнения параллельных заданий. Все эксперименты проводились на сегменте суперкомпьютера MBC-100K [11], состоящем из восьми узлов с установленным программным обеспечением Proxmox VE и четырех узлов со стандартным набором программного обеспечения. В качестве тестовых заданий использовались программы из набора NAS Parallel Benchmarks (NPB). Авторами проведены две серии экспериментов, позволившие оценить:

- влияние установленных на узлы MBC-100K компонентов Proxmox VE при использовании сервиса IaaS. Тесты NPB запускались на вычислительных узлах до установки и активации компонентов Proxmox VE и после;
- влияние виртуальной среды Proxmox VE при использовании сервиса PaaS для параллельных вычислений. Вычисления производились с применением виртуальных машин Proxmox VE с тем же составом программного обеспечения, что и на вычислительных узлах MBC-100K, запущенных через СУППЗ в автоматическом режиме.

При проведении первой серии экспериментов не выявлено существенных изменений во времени выполнения и результатах тестовых заданий NPB (рис. 4).

При выполнении второй серии экспериментов было определено, что время выполнения параллельных программ CG, EP, FT не изменилось. Для остальных программ время выполнения в среде Proxmox VE (сервис PaaS) увеличилось по сравнению с MBC-100K (сервис IaaS) на 2 % для теста BT, на 4 % для тестов MG и SP, на 12 % для теста LU и на 16 % для теста IS (рис. 5). Значительное увеличение времени выполнения программ LU и IS вызвано наличием накладных расходов на коммуникационные операции в виртуальной среде.

Для параллельных программ, чувствительных к накладным расходам, целесообразно использовать сервис IaaS при организации высокопроизводительных вычислений. Однако в ряде довольно распространенных случаев, например в промышленном ренде-

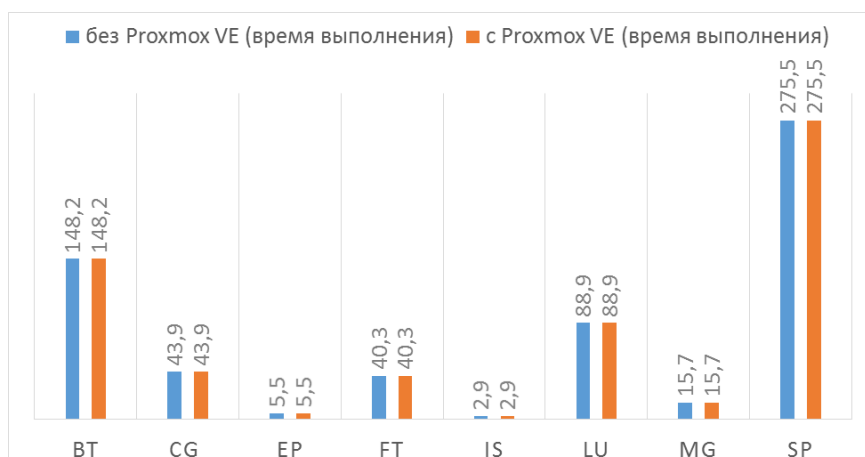


Рис. 4. Оценка влияния установленных на узлы суперкомпьютера компонентов Proxmox VE на время выполнения тестов NPB

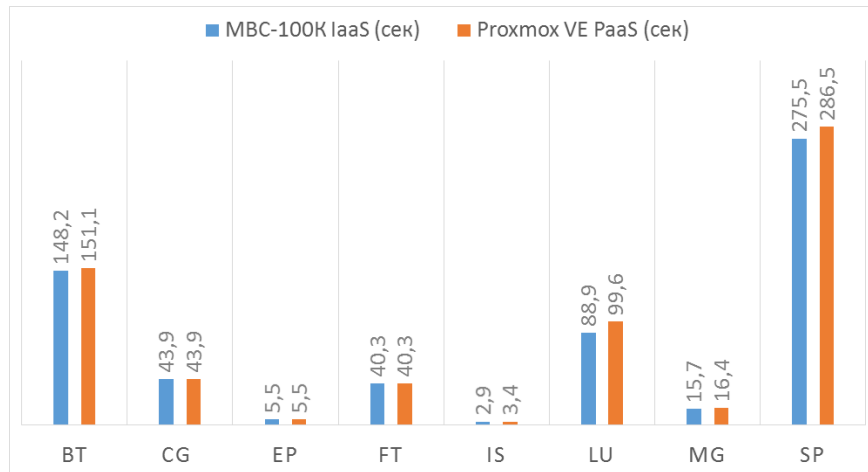


Рис. 5. Оценка влияния виртуальной среды Proxmox VE на время выполнения тестов NPV

ринге [12] или в образовательном процессе при организации учебных занятий по параллельному программированию [13], могут быть построены и использованы сервисы вида PaaS и SaaS. О такой возможности свидетельствует успешное практическое применение авторами предложенных решений при построении защищенной инфраструктуры для решения задач фотореалистичной компьютерной графики [12] и виртуальной учебной лаборатории параллельного программирования [13].

Заключение

Представлен подход к интеграции системы управления заданиями суперкомпьютера и платформы виртуализации, основанный на представлении суперкомпьютерных заданий в виде виртуальных машин или контейнеров. На примере платформы Proxmox VE экспериментально исследовано влияние виртуальной среды на время выполнения параллельных программ. Поскольку для достаточно широкого класса приложений вносимые средой Proxmox накладные расходы являются вполне приемлемыми, можно сделать вывод о возможности применения предложенного подхода для построения облачных сервисов вида PaaS и SaaS в научных суперкомпьютерных центрах коллективного пользования.

Благодарности. Публикация выполнена в рамках государственного задания МСЦ РАН по теме № 0065-2019-0014 “Исследование и разработка методов и средств организации высокопроизводительных вычислений, создания, обработки, хранения и распределения больших данных и цифрового контента в распределенных информационных и вычислительных средах”

Список литературы / References

- [1] Tadviser. Государство. Бизнес. ИТ. Облачные вычисления. 2015. Адрес доступа: [http://www.tadviser.ru/index.php/Статья:Облачные_вычисления_\(мировой_рынок\)](http://www.tadviser.ru/index.php/Статья:Облачные_вычисления_(мировой_рынок)) (дата обращения 21.11.2019).
Tadviser. Government. Business. IT. Cloud computing. Available at: http://tadviser.com/index.php/Cloud_Computing (accessed 21.11.2019)

- [2] Tadviser. Государство. Бизнес. ИТ. Облачные сервисы (рынок России). 2019. Tadviser. Government. Business. IT. Cloud services in the Russian market, 2019. Available at: http://www.tadviser.ru/index.php/Статья:Облачные_сервисы_%28рынок_России%29 (accessed 21.11.2019). (In Russ.)
- [3] **Reuther, A., Byun, C., Arcand, W. et al.** Scalable system scheduling for HPC and big data // J. of Parallel and Distributed Computing. 2018. Vol. 111. P. 76–92. DOI: 10.1016/j.jpdc.2017.06.009.
- [4] **Кудрявцев А.О., Кошелев В.К., Избышев А.О., Аветисян А.И.** Высокопроизводительные вычисления как облачный сервис: ключевые проблемы // Тр. Междунар. конф. “Параллельные вычислительные технологии 2013 (ПаВТ-2013)”. Челябинск: ЮУрГУ, 2013. С. 432–438.
Kudryavtsev, A.O., Koshelev, V.K., Izbyshv, A.O., Avetisyan, A.I. HPC as a cloud service: the key problems // Proc. of the Intern. Conf. “Parallel Computational Technologies (PCT) 2013”. Chelyabinsk: UYUrGU, 2013. P. 432–438. (In Russ.)
- [5] **Аладышев О.С., Баранов А.В., Овсянников А.П. и др.** Методы и средства совмещения потоков заданий от облачных платформ и менеджеров управления ресурсами суперкомпьютера // Программные продукты, системы и алгоритмы. 2018. № 4. С. 54–63.
Aladyshv, O.S., Baranov, A.V., Ovsyannikov, A.P. et al. Methods and tools for combining job streams from cloud platforms and managing of supercomputer resources // Software J.: Theory and Appl. 2018. No. 4. P. 54–63. DOI: 10.15827/2311-6749.29.337. (In Russ.)
- [6] Proxmox Main Page. Available at: https://pve.proxmox.com/wiki/Main_Page (accessed 03.10.2019).
- [7] Система управления прохождением параллельных заданий (СУППЗ). Руководство программиста (пользователя). Parallel jobs management system (SUPPZ). Programmer (user) guide. Available at: <http://www.jssc.ru/wp-content/uploads/2017/06/SUPPZ-user-guide-2016.pdf> (accessed 03.10.2019). (In Russ.)
- [8] Slurm Containers Guide, 2019. Available at: <https://slurm.schedmd.com/containers.html> (accessed 03.10.2019).
- [9] **Баранов А.В., Николаев Д.С.** Использование контейнерной виртуализации в организации высокопроизводительных вычислений // Программные системы: теория и приложения. 2016. № 1(28). С. 117–134.
Baranov, A.V., Nikolaev, D.S. The use of container virtualization in the organization of high-performance computing // Program Systems: Theory and Applications. 2016. No. 1(28). P. 117–134. DOI: 10.25209/2079-3316-2016-7-1-117-134. (In Russ.)
- [10] **Baranov, A.V., Savin, G.I., Shabanov, B.M. et al.** Methods of jobs containerization for supercomputer workload managers // Lobachevskii J. of Mathematics. 2019. Vol. 40, No. 5. P. 525–534. DOI: 10.1134/S1995080219050020.
- [11] Вычислительные ресурсы МСЦ РАН. MVS-100K. Computing resources JSCC RAS. MVS-100K. <http://www.jssc.ru/resources/hpc/#item88> (accessed 03.10.2019). (In Russ.)
- [12] **Шабанов Б.М., Овсянников А.П., Баранов А.В. и др.** Решение задач фотореалистичной компьютерной графики на базе защищенной инфраструктуры суперкомпьютерного центра коллективного пользования // Тр. Междунар. конф. “Суперкомпьютерные дни в России”. М.: Изд-во МГУ, 2017. С. 733–741.
Shabanov, B.M., Ovsyannikov, A.P., Baranov, A.V. et al. Solving problems for photorealistic computer graphics based on the protected infrastructure of a supercomputer collective sharing center // Proc. of the Intern. Conf. “Russian Supercomputing Days”. Moscow: Izd-vo MGU, 2017. P. 733–741. (In Russ.)

- [13] **Аникеев Ф.А., Баранов А.В., Зайцев Ф.С. и др.** Организация виртуальной учебной лаборатории параллельного программирования на базе платформы виртуализации Proxmox VE // Тр. науч.-исслед. ин-та сист. исслед. РАН. 2018. Т. 8, № 6. С. 103–111.
Anikeev, F.A., Baranov, A.V., Zaytsev, F.S. et al. The virtual educational laboratory based on the Proxmox VE virtualization platform // Tr. Nauch.-Issled. In-ta Sist. Issled. RAN. 2018. Vol. 8, No. 1(28). P. 103–111. DOI: 10.25682/NIISI.2018.6.0014. (In Russ.)

Поступила в редакцию 11 октября 2019 г.

HPC cloud services based on the Proxmox VE platform

BARANOV, ANTON V.*, KISELEV, EVGENIY A.

Joint SuperComputer Center RAS — branch of SRISA RAS, Moscow, 119334, Russia

*Corresponding author: Baranov, Anton V., e-mail: antbar@mail.ru

Purpose. The purpose of this paper is to develop methods and technologies for building high-performance computing cloud services in scientific supercomputer centers.

Methodology. To build a cloud environment for high-performance scientific calculations (HPC), the corresponding three-level model and the method of combining flows of supercomputer tasks of various types were applied.

Results. A high-level HPC cloud services technology based on the free Proxmox VE software platform has been developed. The Proxmox VE platform has been integrated with the domestic supercomputer job management system called SUPPZ. Experimental estimates of the overheads introduced in the high-performance computing process by the Proxmox components are obtained.

Findings. An approach to the integration a supercomputer job management system and a virtualization platform is proposed. The presented approach is based on the representation of the supercomputer jobs as virtual machines or containers. Using the Proxmox VE platform as an example, the influence of a virtual environment on the execution time of parallel programs is investigated experimentally. The possibility of applying the proposed approach to building cloud services of the PaaS and SaaS type in scientific supercomputing centers of collective use is substantiated for a class of applications for which the overhead costs introduced by the Proxmox components are acceptable.

Keywords: supercomputer, HPC, cloud service, job management system, scheduling supercomputer jobs, Proxmox.

Cite: Baranov A.V., Kiselev E.A. HPC cloud services based on the Proxmox VE platform // Computational Technologies. 2019. Vol. 24, No. 6. P. 5–12. (In Russ.) DOI: 10.25743/ICT.2019.24.6.002.

Acknowledgements. The research was carried out at JSCC RAS under the State assignment 0065-2019-0014.

Received October 11, 2019