

Опыт ФИЦ ИУ РАН в предоставлении облачных сервисов высокопроизводительных вычислений для задач материаловедения

© 2020 г. С. А. Денисов^{1,§}, К. И. Волович¹, В. А. Кондрашев¹

¹ *Федеральный исследовательский центр «Информатика и управление»
Российской академии наук,
ул. Вавилова, д. 44, корп. 2, Москва, 119333, Россия*

Аннотация. Для решения задач материаловедения в том числе многомасштабного моделирования для синтеза материалов с заданными свойствами в ФИЦ ИУ РАН создана современная цифровая платформа для научных исследований, представляющая собой совокупность центра компетенций, высокопроизводительного вычислительного комплекса и комплекса научных сервисов, предоставляемых исследователям в виде традиционных облачных услуг в режимах программных (SaaS), платформенных (PaaS) и инфраструктурных (IaaS) сервисов, а также с помощью специфических технологий предоставления исследователям научного сервиса как услуги (RaaS, Research as a Service).

Другими примерами научных областей для которых применяются научные сервисы в совокупности с сервисами высокопроизводительных вычислений являются: биомедицинская химия, кристаллография, компьютерная лингвистика, искусственный интеллект.

В статье приводится описание информационно-вычислительной среды ЦКП «Информатика», составляющей основу инструментально-технологической инфраструктуры для макетирования, а также макета системы управления детерминированными научными сервисами цифровой платформы.

В статье представлены результаты экспериментальных исследований, проведенных применительно к алгоритмам передачи и промежуточного хранения исходных данных, сервисам обмена данными при взаимодействии с пользователем платформы, облачным научным сервисам высокопроизводительных вычислений, алгоритмам взаимодействия адаптеров обмена данными при обеспечении взаимодействия между платформой, актуальных при решении задач многомасштабного моделирования для синтеза материалов с заданными свойствами.

Полученные результаты позволяют оценить практические аспекты функционирования цифровой платформы для научных исследований, предназначенной для эффективной организации научных исследований и управления научной приборной базой в интересах широкого круга исследовательских коллективов и промышленных потребителей.

Ключевые слова: научный сервис, консолидированный научный сервис, цифровая платформа, облачные вычисления, сервис-ориентированная архитектура, микросервисы, гибкая интеграция

Введение

Для повышения эффективности проведения экспериментальных исследований в интересах цифровой трансформации [1] в ФИЦ ИУ РАН создана современная цифровая платформа для научных исследований. Она представляет собой совокупность центра компетенций, в котором концентрируются знания в конкретной области, высокопроизводительного вычислительного комплекса и комплекса научных сервисов (аналитических, образовательных, библиотечных, вычислительных, аналитических и др.), которые предоставляют услуги различным сферам деятельности (образованию, науке, коммерции, промышленности, государственным структурам) [2—5].

Особое внимание уделяется исследованиям по многомасштабному моделированию для синтеза материалов с заданными свойствами [6, 7], которые направлены на разработку алгоритмов машинного

Денисов Сергей Анатольевич^{1,§} — ведущий инженер, e-mail: sdenisov@frccsc.ru; **Волович Константин Иосифович**¹ — канд. техн. наук, старший научный сотрудник, e-mail: kvolovich@frccsc.ru; **Кондрашев Вадим Адольфович**¹ — канд. техн. наук, заместитель директора, e-mail: vkondrashev@frccsc.ru

§ Автор для переписки

Статья подготовлена по материалам доклада, представленного на II-й международной конференции «Математическое моделирование в материаловедении электронных компонентов», Москва, 19—21 октября 2020 г. (Денисов С.А., Волович К.И., Кондрашев В.А. Опыт ФИЦ ИУ РАН в предоставлении облачных сервисов высокопроизводительных вычислений для задач материаловедения. М.: МАКС Пресс, 2020. С. 26—29. DOI: 10.29003/m1510.MMMSEC-2020/26-29)

обучения и методов управления большими данными для решения задач квантово-механического моделирования. Использование нейросетевых алгоритмов позволяет с высокой точностью находить решение задач расчета электронной плотности при заданной конфигурации атомов, а также поиска атомной структуры, минимизирующей энергию системы. Выполняются работы в области полупроводникового материаловедения, высокоскоростного взаимодействия твердых тел (внедрение) и моделирования свойств композитных материалов с дефектами.

Для решения выше обозначенных задач материаловедения цифровая платформа предоставляет исследователям высокопроизводительные вычислительные ресурсы в виде традиционных облачных услуг в режимах программных (SaaS), платформенных (PaaS) и инфраструктурных (IaaS) сервисов, а также с помощью специфических технологий предоставления исследователям научного сервиса как услуги (RaaS, *Research as a Service*) [8]. В экосистеме цифровой платформы возможно создание облачных сред для интегрированных систем моделирования ABINIT, VASP, Quantum ESPRESSO, LAMMPS, MEER, MPB, OpenFOAM и др., что позволяет проводить практически любые научные и научно-практические расчеты.

Настоящая работа представляет практические аспекты реализации этих решений на макете системы управления детерминированными научными сервисами цифровой платформы в части обеспечения внешнего интерфейса с пользователями платформы и результаты экспериментальных исследований на макете по следующим направлениям:

- экспериментальное исследование алгоритмов передачи и промежуточного хранения исходных данных для проведения исследований и результатов исследований;
- моделирование функционирования сервиса обмена данными при взаимодействии с пользователем платформы, эксперименты по передаче больших объемов данных;
- экспериментальное исследование функционирования облачных научных сервисов высокопроизводительных вычислений;
- экспериментальное исследование алгоритмов взаимодействия адаптеров обмена данными при обеспечении взаимодействия между платформой и сервисами высокопроизводительных вычислений.

Макетирование системы управления детерминированными научными сервисами проводилось для среды облачных вычислений с применением технологий виртуализации (контейнеризации) в парадигме сервис-ориентированной архитектуры на основе положений технологии «гибкой интеграции» (*Agile Integration*), web-интерфейсов и микросервисов. Для макета использовались средства Центра

коллективного пользования ФИЦ ИУ РАН (ЦКП «Информатика») [9].

Инструментально-технологическая среда для макетирования

Основу ЦКП «Информатика» составляет информационно-вычислительная среда, включающая в себя среду высокопроизводительных вычислений и среду виртуализации, функционирующие на технических средствах ЦОД ФИЦ ИУ РАН [10] (рис. 1).

В состав технических средств ЦОД входят:

- телекоммуникационные средства (коммутаторы, маршрутизаторы, межсетевые экраны) в совокупности образующие телекоммуникационную подсистему ЦОД;
- серверная группировка ЦОД для обработки информации и высокопроизводительных вычислений;
- технические средства системы хранения данных.

Телекоммуникационные средства образуют следующие основные вычислительные сети:

- сеть управления информационно-вычислительной и инженерной инфраструктурой ЦОД — низкоскоростная сеть Ethernet (1 Гбит/с);
- информационная вычислительная сеть — высокоскоростная сеть Ethernet (10/40 Гбит/с);
- сеть системы хранения данных — сеть Fibre Channel (16 Гбит/с);
- высокоскоростная сеть межузлового взаимодействия кластера высокопроизводительных вычислений — сеть InfiniBand (100 Гбит/с).

Серверная группировка ЦОД включает два типа серверов: информационные сервера виртуальной инфраструктуры и высокопроизводительные вычислительные сервера со следующими интегральными характеристиками:

- общее число серверов — 22 шт.;
- общее число центральных процессоров (CPU) — 41 шт.;
- общее число графических ускорителей (GPU) — 30 шт.;
- общее число ядер центральных процессоров (CPU Core) — 728 шт.;
- общий пул оперативной памяти (RAM) — 13 ТБ.

Система хранения данных геораспределенное метространище на основе сети FiberChannel общим объемом чуть более 1 ПБ.

Среда высокопроизводительных вычислений ЦОД (HPC, *High Performance Computing*) включает:

- гибридный высокопроизводительный вычислительный комплекс архитектуры IBM Power9 с GPU Nvidia V100 NVlink (2 узла, 8 GPU);

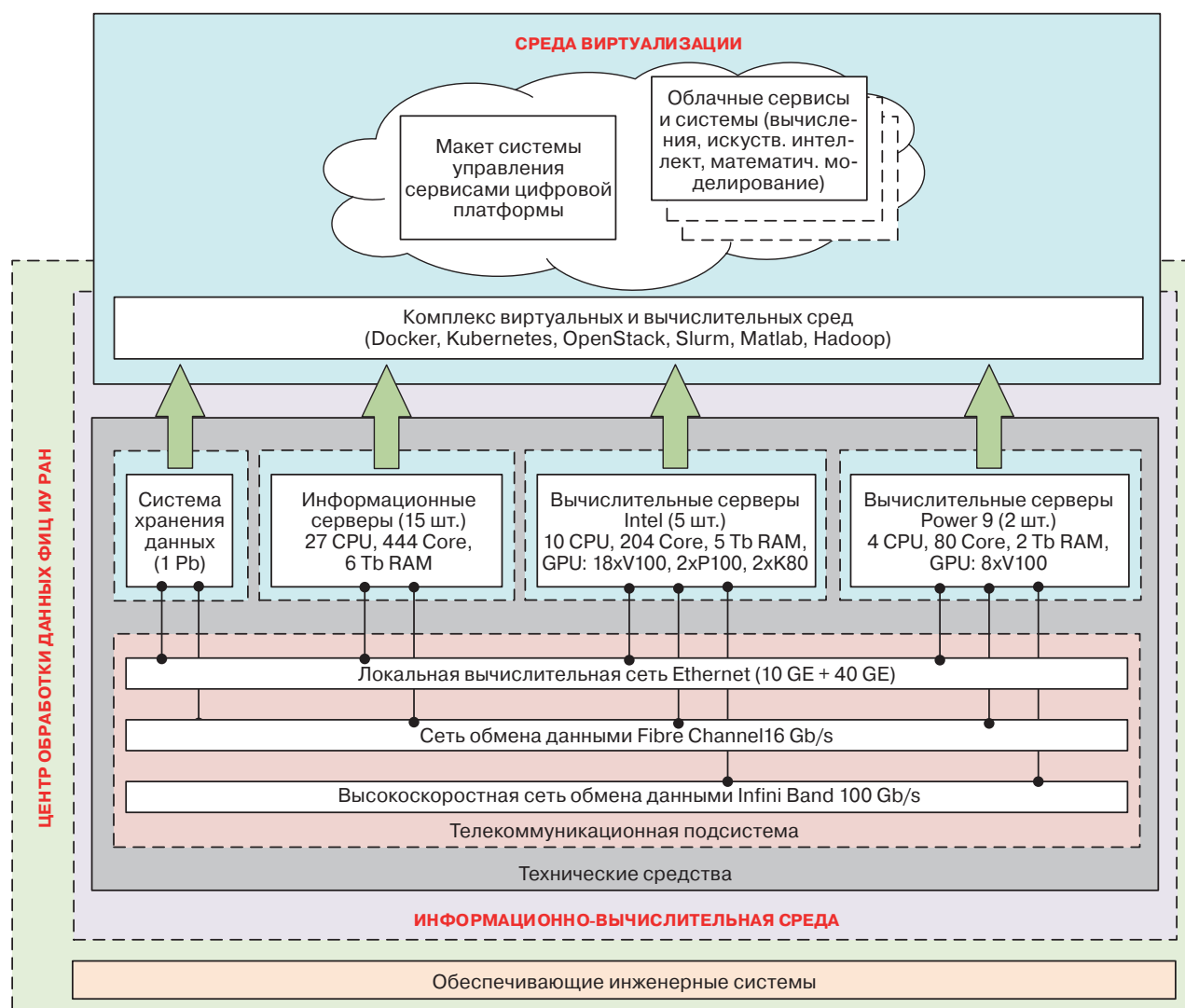


Рис. 1. Инструментально–технологическая среда ЦКП «Информатика»

Fig. 1. Instrumental and technological environment of the Center for Collective Use "Informatics"

– гибридный высокопроизводительный вычислительный комплекс архитектуры Intel с GPU Nvidia V100 (2 узла, 16 GPU);

– вычислительный комплекс (БК) архитектуры Intel с GPU Nvidia P100 и K80.

Комплексы объединены в высокопроизводительный вычислительный кластер под управлением диспетчера задач Slurm и предназначены для решения широкого спектра задач, относящихся к классу HPC. Диспетчер задач Slurm предоставляет пользователям широкие возможности для решения вычислительных задач математического моделирования и искусственного интеллекта посредством запуска вычислительных заданий:

- на одном сервере из состава комплекса;
- на группе серверов, относящихся к одному комплексу;
- на серверах и/или группе серверов, относящихся к разным комплексам.

Среда виртуализации обеспечивает функционирование широкого круга информационно–

вычислительных систем, предоставляет возможность построения экспериментальных стендов и создания макетов информационных облачных систем, в том числе макета системы управления детерминированными научными сервисами цифровой платформы.

В работах [11–13] описаны архитектура, подходы и решения, применяемые при создании макета системы управления детерминированными научными сервисами цифровой платформы (рис. 2). Предлагаемые в этих работах решения и подходы позволяют широкому кругу исследователей и промышленных потребителей воспользоваться достоинствами технологий цифровой экономики в сфере науки.

Интеграционная система является центральным элементом макета и обеспечивает взаимодействие систем макета.

Система ведения каталога научных сервисов является основой предоставления научных сервисов цифровой платформой и предназначена для спецификации научного сервиса.

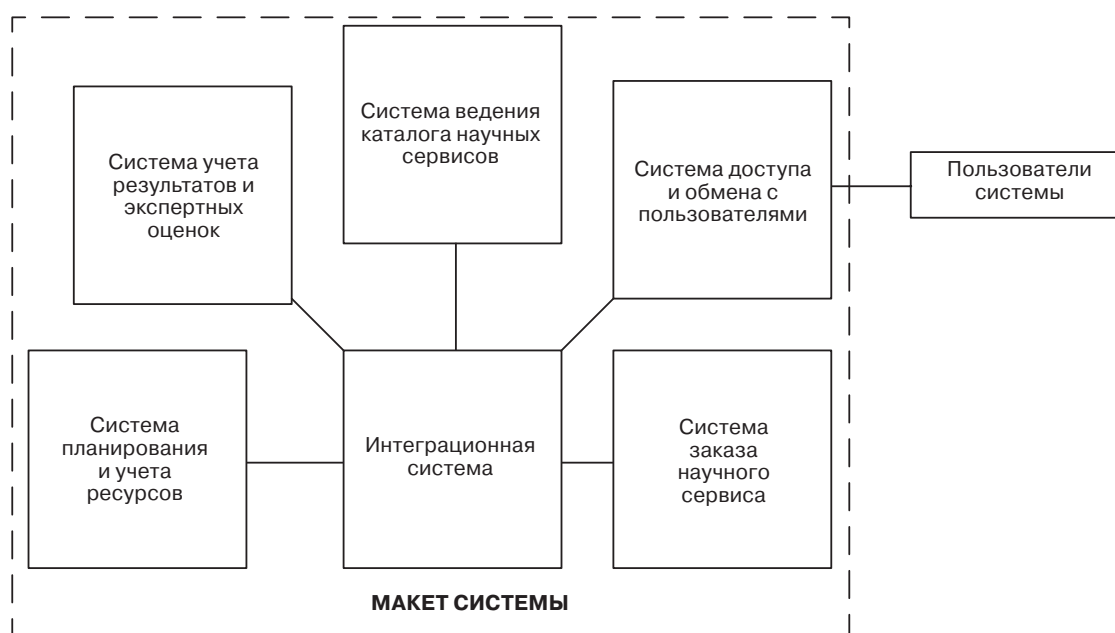


Рис. 2. Взаимосвязь функциональных компонентов макета системы управления сервисами цифровой платформы
Fig. 2. Interrelation of the functional components of the layout of the digital platform service management system

Система заказа научного сервиса предназначена для документирования процессов оказания научной услуги в результате выполнения процессов поиска, выбора и заказа научной услуги во взаимодействии с каталогом научных сервисов, системой планирования и учета ресурсов, а также системой учета результатов и экспертных оценок.

Система планирования и учета ресурсов предназначена для ведения реестра ресурсов (персонала, приборной базы, расходных материалов) и их состояния, а также календаря загрузки и резервирования ресурсов.

Система учета результатов и экспертных оценок предназначена для ведения реестра результатов и накопления оценок предоставленного сервиса в реестре заказов.

Система доступа и обмена с пользователями обеспечивает внешний веб-интерфейс взаимодействия пользователя с цифровой платформой научных сервисов для трех категорий пользователей: заказчика сервиса, поставщика сервиса и оператора (администратора, аналитика) платформы.

Далее в статье представлены результаты экспериментальных исследований алгоритмов взаимодействия пользователя с научными сервисами цифровой платформы по следующим направлениям:

- экспериментальное исследование алгоритмов передачи и промежуточного хранения исходных данных для проведения исследований и результатов исследований;
- моделирование функционирования сервиса обмена данными при взаимодействии с пользователем платформы, эксперименты по передаче больших объемов данных;

- экспериментальное исследование функционирования облачных научных сервисов высокопроизводительных вычислений;

- экспериментальное исследование алгоритмов взаимодействия адаптеров обмена данными при обеспечении взаимодействия между платформой и сервисами высокопроизводительных вычислений.

Экспериментальные исследования функционирования облачных научных сервисов высокопроизводительных вычислений

Экспериментальное исследование алгоритмов передачи и промежуточного хранения исходных данных для проведения исследований и результатов исследований. Передача исходных данных от источника данных к среде обработки данных, расположенной в облачной цифровой платформе, сама по себе является предметом исследований и поиска оптимального алгоритма функционирования.

Алгоритм передачи данных должен учитывать объем необходимых для исследования данных, производительность источника данных и каналов передачи данных, потребность вычислителя в скорости доставки данных, возможность повторного получения данных из источников.

С целью проведения исследований по выбору тех или иных алгоритмов передачи данных были проведены эксперименты по моделированию следующих исходных ситуаций.

Вариант 1. Объем данных, необходимых для вычислений не превосходит возможности хранилища данных вычислительного комплекса. Производительность источника данных и каналов связи

позволяют осуществить передачу данных за приемлемое время до начала эксперимента. Потребность вычислителя в потоке данных удовлетворяется производительностью хранилища данных вычислительного комплекса.

Вариант 2. Объем данных, необходимых для вычислений превосходит возможности хранилища данных вычислительного комплекса. Возможна загрузка части необходимых данных в промежуточное хранилище. Потребность вычислителя в потоке данных удовлетворяется производительностью хранилища данных вычислительного комплекса.

Вариант 3. Объем данных, необходимых для вычислений превосходит возможности хранилища данных вычислительного комплекса. Возможна загрузка часть необходимых данных в промежуточное хранилище. Потребность вычислителя в потоке данных не удовлетворяется производительностью хранилища данных вычислительного комплекса.

Вариант № 1 является тривиальным и для его реализации требуется последовательная загрузка исходных данных из источника данных в хранилище вычислительного комплекса с последующим выполнением расчетной задачи.

В варианте № 2 для организации передачи данных используется промежуточный буфер, в который до начала вычислений загружается массив исходных данных из источника данных, после чего выполняется вычислительная задача. Параллельно с выполнением вычислительной задачи осуществляется передача данных из источника данных в освобождающиеся области промежуточного буфера. В случае опустошения промежуточного буфера выполнение расчетной задачи приостанавливается до заполнения буфера данными из источника данных.

Наиболее сложным с точки зрения организации вычислительного процесса является вариант № 3. В этом случае производительности промежуточного хранилища недостаточно для предоставления вычислительной задаче исходных данных с производительностью, обеспечивающей непрерывное выполнение вычислительной задачи. Требуются дополнительные системотехнические решения, позволяющие внутри существующего вычислительного комплекса создать высокопроизводительное промежуточное хранилище данных.

В эксперименте в качестве высокопроизводительного промежуточного хранилища использовался выделенный в системе хранения данных ЦОД ФИЦ ИУ РАН блок накопителей SSD, обладающих высокой скоростью передачи данных. Такое системотехническое решение позволяет обеспечить увеличение потока данных от хранилища к вычислителю и преодолеть ограничение по производительности источника.

Отметим, что другими методами повышения производительности промежуточного хранилища

данных является использование параллельных файловых систем, объединяющих группу узлов хранения данных и узлов потребителей данных. В качестве примеров таких решения можно привести параллельную файловую систему Lustre или проприетарное решение компании IBM — Elastic Storage.

Эксперименты с передачей и промежуточным хранением массивов данных, предназначенных для решения научных задач показали, что для оптимального обеспечения расчетных задач исходными данными требуются как алгоритмические, так и системотехнические решения, направленные на организацию вычислительного процесса и создание высокопроизводительных компонентов хранения данных на основе открытых либо проприетарных решений.

Моделирование функционирования сервиса обмена данными при взаимодействии с пользователем платформы, эксперименты по передаче больших объемов данных. Моделирование функционирования сервиса обмена данными заключалось в проведении экспериментов по передаче/загрузки пользователем данных из различных web-интерфейсов цифровой платформы, поддерживающих функции передачи/получения данных, в существующую систему хранения данных информационно-вычислительной среды ЦОД ФИЦ ИУ РАН, в которой были выделены три области памяти, имитирующие коллективное, персональное и промежуточное хранилище данных цифровой платформы. Взаимодействие сервиса обмена данными с системой хранения данных осуществлялось по протоколу NFS.

В ходе экспериментов было установлено, что:

1) по управляющим воздействиям пользователей сервисы, отвечающие за работу web-интерфейсов, осуществляют вызов экземпляра сервиса обмена данными и передают ему список загружаемых данных;

2) осуществляется двустороннее взаимодействие сервиса обмена данными с сервисом авторизации; сервис обмена данными отправляет сервису авторизации запросы о допустимости выполнения пользователем операции загрузки/получения данных и получает ответы от сервиса авторизации;

3) сервисом обмена данных осуществляется проверка объема и формат загружаемых данных на соответствие объему и форматам, допустимым для формы загрузки; определяется тип хранилища, в которое будут помещены данные; определяются права доступа других пользователей к загружаемым данным и осуществляется загрузка данных во временное хранилище;

4) осуществляется двустороннее взаимодействие сервиса обмена данными с сервисом антивирусного сканирования. Сервис обмена данными отправляет сервису антивирусного сканирования

запросы о допустимости загрузки данных из временного хранилища и получает ответы от сервиса антивирусного сканирования;

5) сервисом обмена данных осуществляется передача управляющих команд в систему хранения данных на перемещение данных из временного хранилища в коллективное или персональное хранилище.

Эксперименты по передаче и обработке больших данных проводились на гибридном высокопроизводительном комплексе архитектуры IBM Power 9, подключенному к системе хранения данных по протоколу NFS. Эксперименты по передаче больших объемов данных проводились путем генерации случайных наборов данных вне вычислительного комплекса и их передаче в систему хранения данных. Эксперименты показали, что узким местом являются каналы связи, соединяющие источники данных с вычислительным комплексом.

Экспериментальное исследование функционирования облачных научных сервисов высокопроизводительных вычислений. Основной задачей функционирования научных сервисов высокопроизводительных вычислений в облачной среде является предоставление пользователям доступа к высокопроизводительным вычислительным средам (HPC) в виде SaaS, PaaS, IaaS посредством технологий виртуализации, «гибкой интеграции» и web-интерфейсов.

Экспериментальные исследования функционирования научных сервисов высокопроизводительных вычислений макета системы управления детерминированными научными сервисами цифровой платформы заключались в предоставлении пользователям доступа к ресурсам гибридного высокопроизводительного вычислительного комплекса (ГВБК) архитектуры IBM Power 9 в среде высокопроизводительных вычислений ЦОД ФИЦ ИУ РАН.

При проведении экспериментальных исследований была смоделирована работа сервисов, относящихся к типам SaaS и PaaS [14–17]. В ходе экспериментов выполнялись следующие проверки:

1) проверка получения пользователем доступа к сервисам высокопроизводительных вычислений и, как следствие, к интерпретатору командной строки индивидуальной среды исполнения ГВБК; в результате работы сервиса доступа по VPN-сети и системы доступа пользователей к сервисам цифровой платформы пользователь получал доступ к сервисам высокопроизводительных вычислений;

2) проверка обмена данными пользователем с цифровой платформой; в результате работы сервиса обмена данными между пользователями и цифровой платформой пользователь обменивался данными с цифровой платформой;

3) проверка получения пользователем доступа к своим данным в индивидуальной среде исполнения ГВБК; в результате работы адаптера для обмена

данными между сервисами цифровой платформы и сервисами поставщиков услуг пользователь получал доступ к своим данным в индивидуальной среде исполнения ГВБК.

В результате экспериментов по предоставлению облачного сервиса высокопроизводительных вычислений типа PaaS пользователь получал облачный программно-технический ресурс, позволяющий формировать программную среду и решать научную задачу с использованием ресурсов высокопроизводительной вычислительной среды. Опыт ФИЦ ИУ РАН показывает, что сервисы типа PaaS являются наиболее востребованными при решении научных задач, поскольку предоставляют максимальную гибкость при формировании индивидуальной программной среды, что особенно важно для исследовательских работ.

Эксперименты по предоставлению облачных сервисов SaaS проводились в едином комплексе исследований с сервисами PaaS. Для этого на этапе формирования индивидуальной среды исполнения после завершения всех операций по формированию и настройке виртуального контейнера образ этого контейнера помещался в репозиторий вычислительного комплекса и становился доступным для создания новых контейнеров.

В результате пользователь получал возможность создать готовую вычислительную среду из репозитория без дополнительных настроек, что позволяет после создания контейнера сразу переходить к этапу загрузки исходных данных и выполнения приложений.

Сервисы SaaS могут быть востребованы для многократного повторения высокопроизводительных вычислений, не требующих изменения вычислительной среды. Такой подход позволяет экономить ресурсы вычислительного комплекса.

Примерами научных областей для которых применялись сервисы высокопроизводительных вычислений являются: биомедицинская химия, кристаллография, материаловедение, компьютерная лингвистика, искусственный интеллект и другие.

Сервисы высокопроизводительных вычислений для задач биомедицинской химии были использованы для поиска молекулярных структур, обладающих биологической активностью по отношению к мускариновым рецепторам, которые являются важными биомолекулами, ингибиторы которых рассматриваются как лекарственные средства для лечения целого спектра патологических состояний [18], а также для построения универсальной модели предсказания значений IC₅₀ ингибиторов нейраминидазы вируса гриппа произвольного штамма [19].

Сервисы высокопроизводительных вычислений для задач кристаллографии применялись для построения энергетических моделей молекулярного взаимодействия [20].

Для задач вычислительного материаловедения, включая задачи молекулярно-динамического моделирования процесса высокоскоростного взаимодействия частиц [21] и теплопереноса в многослойных структурах проводилось развертывание специфической интегрированной среды VASP на основе базового контейнера с программным обеспечением CUDA.

В задачах анализа больших объемов текстов выбраны две функции, наиболее востребованные у пользователей и требовательные к объему вычислений, а именно семантический поиск и поиск текстовых заимствований [22]. Для решения данной задачи была развернута нетипичная конфигурация — на основе постоянно функционирующего контейнера, обрабатывающего запросы пользователей к web-серверу.

В качестве примера научного исследования с использованием методом искусственного интеллекта можно привести сегментирование трехмерных медицинских изображений, полученных при помощи компьютерной томографии, с применением нейронных сетей глубокого обучения, проводимого участниками магистерских программ [23].

Экспериментальное исследование алгоритмов взаимодействия адаптеров обмена данными при обеспечении взаимодействия между платформой и сервисами высокопроизводительных вычислений. Адаптеры обмена данными, обеспечивающие взаимодействие пользователя и вычислительного комплекса представляют собой сервисы цифровой платформы, использующие единую службу аутентификации платформы (SSO) для получения доступа к централизованному хранилищу данных. Алгоритм функционирования службы-адаптера предполагает выполнение следующих шагов:

- определение прав пользователя по результатам аутентификации в домене SSO;
- выделение области данных в хранилище данных, которая ассоциирована с пользователем;
- передача данных по инициативе пользователя, либо по инициативе вычислительного комплекса (загрузка данных из источника).

В случае передачи данных по инициативе пользователя:

- предоставление доступа к интерфейсу передачи данных. В качестве протокола передачи данных использовались протоколы SFTP (передача на основе протокола SSH) и HTTPS;
- прием данных от пользователя (источника) и помещение их в хранилище;

В случае передачи данных по инициативе вычислительного комплекса (загрузка данных из источника):

- получение доступа к интерфейсу передачи данных. В качестве протокола передачи данных использовались протокол HTTP, при необходимости с использованием аутентификации;

- загрузка данных от источника и помещение их в хранилище.

Отметим, что облачная среда вычислительного комплекса позволяет создавать внутренние сервисы платформы на базе тех же принципов, что и сервисы, предназначенные для обслуживания пользователей. В эксперименте адаптер обмена данными представлял собой виртуальный контейнер на базе технологии docker, разворачиваемый в облачной среде. Для каждой пользовательской сессии создавался отдельный контейнер, выполняющий описанный выше алгоритм. Таким образом, адаптер передачи данных предоставлял пользователю облачный сервис типа SaaS.

Среда исполнения адаптера, основанная на сценариях языка Python, которые обеспечивают выполнение аутентификации (через SSO), работу по протоколам SFTP, HTTP/HTTPS, взаимодействие с хранилищем данных.

В случае необходимости может быть использован любой другой контейнер, выполняющий другие внутренние алгоритмы, однако, предоставляющий унифицированный интерфейс для пользователя и файлового хранилища.

Эксперименты показали, что такая модульность особенно удобна при взаимодействии с разнотипными источниками данных, интерактивными пользователями и автоматизированными системами, которые могут выступать в роли пользователей высокопроизводительного вычислительного комплекса.

Таким образом, при создании адаптеров обмена данными поддерживается множественность алгоритмов доступа и взаимодействия источника данных и хранилища данных высокопроизводительного комплекса при сохранении интерфейсов взаимодействия.

Выполнению этих условий способствует реализация адаптера данных как облачного сервиса цифровой платформы типа SaaS.

Заключение

В статье представлены результаты экспериментальных исследований, проведенные применительно к алгоритмам передачи и промежуточного хранения исходных данных, сервисам обмена данными при взаимодействии с пользователем платформы, облачным научным сервисам высокопроизводительных вычислений, алгоритмам взаимодействия адаптеров обмена данными при обеспечении взаимодействия между платформой, актуальных при решении задач многомасштабного моделирования для синтеза материалов с заданными свойствами.

Экспериментальные исследования, продемонстрировали следующие результаты:

- обмен данными касательно сведений о времени начала выполнения услуги, исполняемой команде

и исходных данных, о ходе и результате выполнения услуги целесообразно реализовывать с применением технологий web-интерфейсов (web-API в стиле REST или gRPC). Обмен данными, содержащими исполняемую команду, исходные данные, результат выполнения во избежание чрезмерной нагрузки на сервис целесообразно реализовывать посредством механизма сетевых виртуальных устройств хранения данных в облачной среде цифровой платформы;

– сервисы высокопроизводительных вычислений типа PaaS являются наиболее востребованными при решении научных задач, поскольку предоставляют максимальную гибкость при формировании индивидуальной программной среды, что особенно важно для исследовательских работ;

– подход, при котором для каждой пользовательской сессии адаптер обмена данными реализуется в виде виртуального контейнера, является наиболее применимым при взаимодействии с разнотипными источниками данных, поскольку поддерживается множественность алгоритмов доступа и взаимодействия источника данных и хранилища данных высокопроизводительного комплекса.

Полученные результаты позволили оценить практические аспекты функционирования цифровой платформы для научных исследований, предназначенной для эффективной организации научных исследований и управления научной приборной базой в интересах широкого круга исследовательских коллективов и промышленных потребителей в том числе о области материаловедения.

Библиографический список

1. Зацаринный А. А., Киселев Э. В., Козлов С. В., Колин К. К. Информационное пространство цифровой экономики России. Концептуальные основы и проблемы формирования / под общ. ред. А. А. Зацаринного. М.: ФИЦ ИУ РАН; ООО «НИПКЦ Восход-А», 2018. 236 с.
2. Зацаринный А. А., Кондрашев В. А., Сучков А. П. Система научных сервисов как актуальный компонент научных исследований // Системы и средства информатики. 2019. Т. 29, № 1. С. 25—40. DOI: 10.14357/08696527190103
3. Zatsarinny A. A., Kondrashev V. A., Sorokin A. A. Approaches to the organization of the computing process of a hybrid high-performance computing cluster in the digital platform environment // CEUR Workshop Proceedings: ITHPC 2019 – Short Paper Proc. 5th International Conference on Information Technologies and High-Performance Computing, Khabarovsk, 16–19 September 2019. V. 2426. Khabarovsk, 2019. P. 12—16.
4. Зацаринный А. А., Кондрашев В. А., Сорокин А. А., Денисов С. А. Методы консолидации научных сервисов // Известия высших учебных заведений. Материалы электронной техники. 2019. Т. 22, № 4. С. 302—307. DOI: 10.17073/1609-3577-2019-4-302-307
5. Волович К. И., Денисов С. А. Основные научно-технические проблемы применения гибридных высокопроизводительных вычислительных комплексов в материаловедении // Известия высших учебных заведений. Материалы электронной техники. 2019. Т. 22, № 4. С. 262—267. DOI: 10.17073/1609-3577-2019-4-262-267
6. Абгарян К. К. Многомасштабное моделирование работы многоуровневых элементов памяти, применяемых для создания нейроморфных сетей // Математическое моделирование в материаловедении электронных компонентов: материалы I международной конференции, Москва, 21–23 октября 2019 г. М.: МАКС Пресс, 2019. С. 53—55. DOI: 10.29003/m682.MMMSEC-2019
7. Абгарян К. К. Информационная технология построения многомасштабных моделей в задачах вычислительного материаловедения // Системы высокой доступности. 2018. Т. 14, № 2. С. 9—15.
8. Кондрашев В. А., Волович К. И. Управление сервисами цифровой платформы на примере услуги высокопроизводительных вычислений // Математическое моделирование и информационные технологии в инженерных и бизнес-приложениях: материалы международной научной конференции, Воронеж, 3–6 сентября 2018 г., Воронеж: Воронежский государственный университет, 2018. С. 217—223.
9. Положение о Центре коллективного пользования «Информатика». URL: <https://www.frccsc.ru/ckp> (дата обращения: 23.06.2020).
10. Зацаринный А. А., Гаранин А. И., Кондрашев В. А., Волович К. И., Мальковский С. И. Оценка надежности гибридного высокопроизводительного вычислительного комплекса при решении научных задач // Системы и средства информатики. 2019. Т. 29, № 2. С. 135—147. DOI: 10.14357/08696527190212
11. Кондрашев В. А. Архитектура системы предоставления сервисов цифровой платформы для научных исследований // Системы и средства информатики. 2018. Т. 28, № 3. С. 131—140. DOI: 10.14357/08696527180310
12. Zatsarinny A. A., Kondrashev V. A., Sorokin A. A., Denisov S. A. Scientific services consolidation methods // Russian Microelectronics. 2020. V. 49, N 8. P. 612—616. DOI: 10.1134/S1063739720080144
13. Зацаринный А. А., Кондрашев В. А., Сорокин А. А. Алгоритмы управления сервис-ориентированными процессами детерминированных научных сервисов в гибридных вычислительных средах цифровых платформ // Системы высокой доступности. 2020. Т. 16, № 3. С. 5—17. DOI: 10.18127/j20729472-202003-01
14. Волович К. И. Некоторые системотехнические вопросы предоставления вычислительных ресурсов для научных исследований в гибридной высокопроизводительной облачной среде // Системы и средства информатики. 2018. Т. 28, № 4. С. 98—109. DOI: 10.14357/08696527180410
15. Волович К. И., Денисов С. А., Мальковский С. И. Формирование индивидуальной среды моделирования в гибридном высокопроизводительном вычислительном комплексе // Математическое моделирование в материаловедении электронных компонентов: материалы I международной конференции, Москва, 21–23 октября 2019 г. М.: МАКС Пресс, 2019. С. 21—24. DOI: 10.29003/m682.MMMSEC-2019
16. Волович К. И., Денисов С. А. Подходы к организации вычислительного процесса гибридного высокопроизводительного вычислительного комплекса для решения задач материаловедения // Информатика: проблемы, методы, технологии: материалы XX Международной научно-методической конференции, Воронеж, 13–14 февраля 2020 г. Воронеж: “Научно-исследовательские публикации” (ООО “Вэлборн”), 2020. С. 368—377.
17. Volovich K. I., Denisov S. A., Shabanov A. P., Malkovsky S. I. Aspects of the assessment of the quality of loading hybrid high-performance computing cluster // CEUR Workshop Proceedings: ITHPC 2019 – Short Paper Proc. 5th International Conference on Information Technologies and High-Performance Computing, Khabarovsk, 16–19 September 2019. V. 2426. Khabarovsk, 2019. P. 7—11.
18. Микурова А. В., Скворцов В. С., Раевский О. А. Компьютерная оценка селективности ингибирования мускариновых рецепторов M1–M4 // Biomedical Chemistry: Research and Methods. 2018. Т. 1, № 3. С. e00072. DOI: 10.18097/BMCRM00072
19. Микурова А. В., Скворцов В. С. Создание обобщенной модели предсказания ингибирования нейраминидазы вируса гриппа различных штаммов // Биомедицинская химия. 2018. Т. 64, № 3. С. 247—252. DOI: 10.18097/PBMC20186403247
20. Eistrikh-Heller P. A., Rubinsky S. V., Prokofev I. I., Gabdulkhakov A. G., Mironov A. S., Lashkov A. A. X-ray structure and molecular dynamics study of uridine phosphorylase from vibrio cholerae in complex with 2,2'-anhydrouridine // Crystallography Reports. 2020. V. 65, N 2. P. 269—277. DOI: 10.1134/S1063774520020066
21. Абгарян К. К., Ревизников Д. Л., Журавлев А. А. Параллельная обработка данных в задачах компьютерного моделирования высокоскоростного взаимодействия твердых тел //

Материалы XX Юбилейной Международной конференции по вычислительной механике и современным прикладным системам (ВМСППС'2017), Алушта, 24–31 мая 2017 г. Алушта: Изд-во МАИ–Принт, 2017. С. 27–28.

22. Ядринцев В. В., Клубина К. В., Тихомиров И. А., Гершельман А. Ф. Выбор серверного решения для цифровой платформы поиска и анализа текстов // Системы и средства информатики. 2018. Т. 28, № 3. С. 26–38. DOI: 10.14357/08696527180302

23. Аветисян М. С. Сегментирование объемных медицинских изображений с использованием сверточных нейронных сетей // Магистерская диссертация. М.: МГУ им. М.В. Ломоносова, 2018. С. 58. URL: <http://synthesis.ipi.ac.ru/synthesis/student/BigData/master-thesis/2018%20Аветисян.pdf> (дата обращения: 30.06.2020).

Работа выполнена при частичной поддержке РФФИ (проекты № 18–29–03100 и № 18–29–03091) с использованием ресурсов гибридного высокопроизводительного вычислительного комплекса Центра коллективного пользования «Информатика» ФИЦ ИУ РАН.

Статья поступила в редакцию 30 ноября 2020 г.

Izvestiya vuzov. Materialy elektronnoi tekhniki = Materials of Electronics Engineering. 2020, vol. 23, no. 4, pp. 311–320. DOI: 10.17073/1609-3577-2020-4-311-320

Experience of FRC CSC RAS in providing high–performance computing cloud services for materials science problems

S. A. Denisov^{1,§}, K. I. Volovich¹, V. A. Kondrashev¹

¹ *Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44–2 Vavilov Str., Moscow 119333, Russia*

Abstract. To solve the problems of materials science, including multiscale modeling for the synthesis of materials with specified properties, a modern digital platform for scientific research has been created at the FRC CSC RAS. The digital platform is a combination of a competence center, a high–performance computing complex and a set of scientific services that are provided to researchers in the form of traditional cloud services in software (SaaS), platform (PaaS) and infrastructure (IaaS) services, as well as using specific technologies for providing researchers scientific service as a service (RaaS, Research as a Service).

Other examples of scientific fields for which scientific services are used in conjunction with high–performance computing services are: biomedical chemistry, crystallography, computational linguistics, artificial intelligence.

The article describes the information and computing environment of the sharing research facilities Center for Collective Use “Informatics”, which forms the basis of the instrumental and technological infrastructure for prototyping, as well as the layout of the control system for deterministic scientific services of the digital platform.

The article presents the results of experimental studies carried out in relation to algorithms for the transfer and intermediate storage of initial data, data exchange services when interacting with a platform user, cloud scientific high–performance computing services, algorithms for the interaction of data exchange adapters when ensuring interaction between the platform, which are relevant in solving problems of multiscale modeling for the synthesis of materials with desired properties.

The results obtained allow us to evaluate the practical aspects of the functioning of a digital platform for scientific research, designed for the effective organization of scientific research and management of the scientific instrument base in the interests of a wide range of research teams and industrial consumers.

Keywords: scientific service, consolidated scientific service, digital platform, cloud computing, service oriented architecture, microservices, flexible integration

References

1. Zatsarinnyy A. A., Kiselev E. V., Kozlov S. V., Kolin K. K. Information space of the digital economy of Russia. Conceptual foundations and problems of formation. Moscow: OOO “NIPKTS Voskhod–A”, 2018. 236 p. (In Russ.)
2. Zatsarinnyy A. A., Kondrashev V. A., Suchkov A. P. The system of scientific services as a relevant component of scientific research. *Systems and Means of Informatics*. 2019, vol. 29, iss. 1, pp. 25–40. (In Russ.). DOI: 10.14357/08696527190103
3. Zatsarinnyy A. A., Kondrashev V. A., Sorokin A. A. Approaches to the organization of the computing process of a hybrid

high–performance computing cluster in the digital platform environment. *CEUR Workshop Proc.: 5th International Conference on Information Technologies and High–Performance Computing, ITHPC 2019; Computing Center of Far East Branch of the Russian Academy of Science. Khabarovsk, 16–19 September 2019, vol. 2426*. Khabarovsk, 2019. Pp. 12–16.

4. Zatsarinnyy A. A., Kondrashev V. A., Sorokin A. A., Denisov S. A. Scientific services consolidation methods. *Izvestiya Vysshikh Uchebnykh Zavedenii. Materialy Elektronnoi Tekhniki = Materials of Electronics Engineering*. 2019, vol. 22, no. 4, pp. 302–307. (In Russ.). DOI: 10.17073/1609-3577-2019-4-302-307

5. Volovich K. I., Denisov S. A. The main scientific and technical problems of using hybrid HPC clusters in materials science. *Izvestiya Vysshikh Uchebnykh Zavedenii. Materialy Elektronnoi Tekhniki = Materials of Electronics Engineering*. 2019, vol. 22, no. 4, pp. 262–267. (In Russ.). DOI: 10.17073/1609-3577-2019-4-262-267

6. Abgaryan K. K. Multiscale modeling of the operation of multilevel memory elements used to create neuromorphic networks. *Mathematical Modeling in Materials Science of Electronic Compo-*

Information about authors:

Sergey A. Denisov^{1,§} — Lead Engineer, e-mail: sdenisov@frccsc.ru; **Konstantin I. Volovich**¹ — Cand. Sci. (Eng.), Senior Researcher, e-mail: kvolovich@frccsc.ru; **Vadim A. Kondrashev**¹ — Cand. Sci. (Eng.), Deputy Director, e-mail: vkondrashev@frccsc.ru

§ Corresponding author

nent: *Proc. I International Conference. Moscow, October 21–23, 2019*. Moscow: MAKS Press, 2019. Pp. 53–55. (In Russ.). DOI: 10.29003/m682.MMMSEC-2019

7. Abgaryan K. K. Information technology of the construction of multi-scale models in problems of computational materials science. *Highly Available Systems*. 2018, vol. 14, no. 2, pp. 9–15. (In Russ.)

8. Kondrashev V. A., Volovich K. I. Digital platform service management on the example of high-performance computing services. *Mathematical Modeling and Information Technologies in Engineering and Business Applications: Proc. International Scientific Conference. Voronezh, September 3–6, 2018*. Voronezh: Voronezhskiy gosudarstvennyy universitet, 2018. Pp. 217–223. (In Russ.)

9. Regulations on the Center for Collective Use “Informatics”. (In Russ.). URL: <https://www.frcsc.ru/ckp> (accessed: 23.06.2020).

10. Zatsarinny A. A., Garanin A. I., Kondrashev V. A., Volovich K. I., Malkovsky S. I. Evaluation of reliability of the hybrid high-performance computing complex in solution of scientific problems. *Systems and Means of Informatics*. 2019, vol. 29, iss. 2, pp. 135–147. (In Russ.). DOI: 10.14357/08696527190212

11. Kondrashev V. A. Architecture of the service delivery system for the research services digital platform. *Systems and Means of Informatics*. 2018, vol. 28, iss. 3, pp. 131–140. (In Russ.). DOI: 10.14357/08696527180310

12. Zatsarinny A. A., Kondrashev V. A., Sorokin A. A., Denisov S. A. Scientific services consolidation methods. *Russian Microelectronics*. 2020, vol. 49, no. 8, pp. 612–616. DOI: 10.1134/S1063739720080144

13. Zatsarinny A. A., Kondrashev V. A., Sorokin A. A. Algorithms for managing service-oriented processes of determinate research services in hybrid computing environments of digital platforms. *Highly Available Systems*. 2020, vol. 16, no. 3, pp. 5–17. (In Russ.). DOI: 10.18127/j20729472-202003-01

14. Volovich K. I. Organization of calculations in a hybrid high-performance computing cluster for parallel execution of heterogeneous tasks. *Systems and Means of Informatics*. 2018, vol. 28, iss. 4, pp. 98–109. (In Russ.). DOI: 10.14357/08696527180410

15. Volovich K. I., Denisov S. A., Malkovsky S. I. Formation of an individual modeling environment in a hybrid high-performance computer system. *Mathematical Modeling in Materials Science of Electronic Component: Proc. I International Conference. Moscow, October 21–23, 2019*. Moscow: MAKS Press, 2019. Pp. 21–24. (In Russ.). DOI: 10.29003/m682.MMMSEC-2019

16. Volovich K. I., Denisov S. A. Approaches to organizing the computational process of a hybrid high-performance computing complex for solving problems of materials science. *Informatics: Problems, Methods, Technologies: Proc. XX International Scientific and Methodological Conference. Voronezh, February 13–14, 2020*. Voronezh: “Nauchno-issledovatel’skiye publikatsii” (OOO “Velborn”), 2020. Pp. 368–377. (In Russ.)

17. Volovich K. I., Denisov S. A., Shabanov A. P., Malkovsky S. I. Aspects of the assessment of the quality of loading hybrid high-performance computing cluster. *CEUR Workshop Proc.: 5th International Conference on Information Technologies and High-Performance Computing, ITHPC 2019; Computing Center of Far East Branch of the Russian Academy of Science. Khabarovsk, 16–19 September 2019, vol. 2426*. Khabarovsk, 2019. Pp. 7–11.

18. Mikurova A. V., Skvortsov V. S., Raevsky O. A. Computational evaluation of selectivity of inhibition of muscarinic receptors M1–M4. *Biomedical Chemistry: Research and Methods*. 2018, vol. 1, no. 3, p. e00072. (In Russ.). DOI: 10.18097/BMCRM00072

19. Mikurova A. V., Skvortsov V. S. A generalized prediction model of inhibition of neuraminidase of influenza virus of various strains. *Biochemistry (Moscow), Supplement Series B: Biomedical Chemistry*. 2018, vol. 12, no. 4, pp. 322–329. DOI: 10.1134/S1990750818040054

20. Eistrikh–Heller P. A., Rubinsky S. V., Prokofev I. I., Gabdulkhakov A. G., Mironov A. S., Lashkov A. A. X-ray structure and molecular dynamics study of uridine phosphorylase from vibrio cholerae in complex with 2,2′-anhydrouridine. *Crystallography Reports*. 2020, vol. 65, no. 2, pp. 269–277. DOI: 10.1134/S1063774520020066

21. Abgaryan K. K., Reviznikov D. L., Zhuravlev A. A. Parallel data processing in problems of computer simulation of high-speed interaction of solid bodies. *Proc. XX Anniversary International Conference on Computational Mechanics and Modern Applied Systems. Alushta, May 24–31, 2017*. Alushta: Izdatel’stvo MAI–Print, 2017. Pp. 27–28. (In Russ.)

22. Yadrintsev V., Klyubina K., Tikhomirov I., Gershel’man A. Choosing server solution for digital platform of text search and analysis. *Systems and Means of Informatics*. 2018, vol. 28, iss. 3, pp. 26–38. (In Russ.). DOI: 10.14357/08696527180302

23. Avetisyan M. S. *Segmentation of volumetric medical images using convolutional neural networks*. Master’s Thesis. Moscow: MGU im. M.V. Lomonosov, 2018. P. 58. (In Russ.). URL: <http://synthesis.ipi.ac.ru/synthesis/student/BigData/master-thesis/2018%20Аветисян.pdf> (accessed: 30.06.2020).

Acknowledgments

This work was supported in part by the Russian Foundation for Basic Research (project No. 18–29–03100 and No. 18–29–03091) using the resources of the hybrid high-performance computing complex of the of the Center for Collective Use “Informatics” of the Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences.

Received November 30, 2020

* * *