Д.В. Беляков[1], А.С. Воронцов[1], Е.А. Дружинин[2], М.И. Зуев[1], В.В. Кореньков[1], Ю.М. Мигаль[2], А.А. Мошкин[3], Д.В. Подгайный[1], Т.А. Стриж[1], О.И. Стрельцова[1]

[1] Лаборатория информационных технологий им. М.Г. Мещерякова ОИЯИ
[2] ЗАО "РСК Технологии"
[3] Лаборатория физики высоких энергий им. В.И. Векслера и А.М. Балдина ОИЯИ

На конкурс работ ОИЯИ по разделу научно-методических и научно-технических исследований выдвигается цикл работ "**Гиперконвергентный суперкомпьютер «Говорун» для реализации научной программы ОИЯИ**".

Создание суперкомпьютера «Говорун» в ОИЯИ является важным технологическим достижением и имеет большое значение для реализации научной программы и международного сотрудничества ОИЯИ.

Суперкомпьютер «Говорун» был создан в 2018 году на основе опыта, накопленного при эксплуатации гетерогенного кластера HybriLIT, входящего в состав Многофункционального информационно–вычислительного комплекса ЛИТ ОИЯИ [1]. HybriLIT показал свою востребованность при решении задач КХД на решетках, радиационной биологии, в прикладных исследованиях и др. [2] Постоянный рост числа пользователей и расширение круга решаемых задач потребовали не просто существенно нарастить вычислительные возможности кластера, а разработать и внедрить новые технологии, что привело к созданию новой вычислительной системы – суперкомпьютера «Говорун». СК «Говорун» создавался как высокопроизводительная масштабируемая система с жидкостным охлаждением, обладающая гиперконвергентной и программно-определяемой архитектурой. В текущую конфигурацию СК «Говорун» входят вычислительные модули, содержащие GPU и CPU компоненты, а также иерархическая система обработки и хранения данных [3]. Суммарная пиковая производительность суперкомпьютера «Говорун» составляет 1,1 Пфлопс для расчетов с двойной точностью (2,2 Пфлопс для расчетов с одинарной точностью) и скоростью чтения/записи 300 Гб/с для иерархической системы обработки и хранения данных.

Для CPU – компоненты суперкомпьютера была выбрана технология прямого жидкостного охлаждения компании ЗАО "РСК Технологии", являющейся ведущим в России разработчиком и интегратором «полного цикла» суперкомпьютерных решений и обладающей целым рядом собственных инновационных разработок [4, 5]. Благодаря внедрению этих технологий для СК «Говорун» удалось достичь рекордной плотности размещения вычислительных узлов на шкаф (153 узла против 25 узлов для воздушного охлаждения), а работа в режиме охлаждения «горячей водой» позволила использовать круглогодичный режим free cooling (24x7x365). Помимо высокой энергоэффективности, такой подход позволил существенно упростить инфраструктуру суперкомпьютерного центра – система охлаждения СК «Говорун» создана, используя только сухие градирни, охлаждающие жидкость при помощи окружающего воздуха. За счет применения жидкостного охлаждения среднегодовой показатель PUE системы, отражающий уровень эффективности использования электроэнергии, составляет менее чем 1,06. То есть на охлаждение расходуется менее 6% всего потребляемого СК «Говорун» электричества, что является выдающимся результатом для HPC-индустрии. Построенная система является первой в мире системой со 100% жидкостным охлаждением, т.е. жидкостным образом охлаждаются все компоненты – вычислительные узлы, сетевые коммутаторы и система хранения данных.

Другой технологией, положенной в основу СК «Говорун», является гиперконвергентный подход к построению вычислительного комплекса, позволяющий создавать вычислительные среды, программно-аппаратная конфигурация которых оптимизирована для конкретных задач пользователей, без изменения аппаратуры самих вычислительных узлов. Гиперконвергентность позволяет «оркестрировать» вычислительными ресурсами и элементами хранения данных и создавать, используя программное обеспечение РСК БазИС, вычислительные системы, конфигурации которых создаются по требованию, с учетом потребностей пользовательских приложений. Под термином «оркестрация» подразумевается, с одной стороны, логическая дезинтеграция вычислительного узла на отдельные компоненты, такие как вычислительные ядра, элементы хранены данных (SSD накопители) с последующим их объединением в конфигурацию. Таким образом, вычислительные элементы (CPU-ядра и графические ускорители) и элементы хранения данных (SSD диски) образуют независимые наборы ресурсов (пулы). Благодаря оркестрации пользователь может под свою задачу аллоцировать необходимое число и тип вычислительных узлов (в том числе необходимое число графических ускорителей), необходимый объем и тип систем хранения данных, а также автоматически настроить необходимое ПО, в том числе параллельные файловые системы. После завершения задачи вычислительные ядра и элементы хранения возвращаются в соответствующие пулы и готовы к следующему использованию. Это свойство позволяет эффективно решать пользовательские задачи разных типов, повысить уровень конфиденциальности работы с данными, избежать системных ошибок, возникающих при пересечении ресурсов для различных пользовательских задач. Реализованная на гиперконвергентных узлах первой очереди система хранения данных по требованию (“storage-on-demand”) под управлением файловой системы Lustre позволила суперкомпьютеру «Говорун» занять 9-е место в мировом рейтинге IO500 (июнь 2018 г.) для систем хранения данных HPC-класса.

Благодаря гиперконвергентности СК «Говорун» обладает гибкой архитектурой, позволяющей создавать программно-конфигурируемые HPC-подсистемы, что качественно отличает его от других суперкомпьютеров, обладающих, как правило, «жесткой» архитектурой и предназначенных для эффективного решения узкоспециализированных классов задач [6].

Опыт эксплуатации первой очереди СК «Говорун» выявил необходимость не только в наращивании вычислительных ресурсов, что определилось его востребованностью для решения задач ОИЯИ и ростом числа пользователей, но и потребность создания инструментов для работы с Большими данными, прежде всего, для мегапроекта NICA [7, 8]. В связи с этим на СК «Говорун» была разработана и внедрена иерархическая система обработки и хранения данных с программно-определяемой архитектурой. По скорости доступа к данным система разделена на уровни: очень горячие данные – наиболее востребованные данные, к которым в настоящий момент требуется обеспечить самый быстрый доступ, горячие данные и теплые данные. Каждый уровень разработанной системы хранения данных может использоваться как самостоятельно, так и в составе рабочих процессов обработки данных. В настоящий момент на СК «Говорун» в качестве слоя очень горячих данных осуществляется внедрение новейшей технологии DAOS (Distributed Asynchronous Object Storage) для обработки больших данных, показавшей свою перспективность для задач глубокого обучения и для работы квантовых симуляторов при эмуляции большего числа кубитов. За высокоскоростную систему обработки и хранения данных СК «Говорун» получил престижную премию Russian DC Awards 2020 в номинации «Лучшее ИТ-решение для центров обработки данных».

Задачи массовой генерации и реконструкции данных эксперимента MPD NICA активно используют иерархическую систему обработки и хранения данных СК «Говорун» [9]. При этом на разных этапах рабочих процессов возникает потребность в разной скорости доступа к данным, например, для задач долговременного хранения скорость доступа не является важным фактором, а для задач реконструкции – играет существенную роль. Также для ряда задач MPD возникла потребность в большом объеме оперативной памяти, что привело к необходимости включения в архитектуру суперкомпьютера гиперконвергентных узлов с большим объемом памяти. Таким образом, методологически, для обеспечения всех рабочих процессов для задач мегапроекта NICA на СК «Говорун» создана система, сочетающая в себе как вычислительные архитектуры различных типов, так и развитую иерархическую систему обработки и хранения данных. Вычислительные ресурсы и иерархическая система обработки и хранения данных СК «Говорун» были интегрированы на базе платформы DIRAC в распределенную гетерогенную среду, включающую в себя ресурсы ОИЯИ и стран-участниц [10]. Практика использования различных вычислительных ресурсов ОИЯИ и других институтов коллаборации MPD показала, что на данный момент наиболее эффективным является использование ресурсов именно СК «Говорун» [11].

Реализация перечисленных технологий на СК «Говорун» позволила выполнить ряд сложных ресурсоемких расчетов в области решеточной квантовой хромодинамики для исследования свойств адронной материи при высоких плотностях энергии и барионном заряде и в присутствии сверхмаксимальных электромагнитных полей, качественно повысить эффективность моделирования динамики столкновений релятивистских тяжелых ионов, ускорить процесс генерации и реконструкции событий для проведения экспериментов в рамках реализации мегапроекта NICA, провести расчеты радиационной безопасности экспериментальных установок ОИЯИ и повысить эффективность решения прикладных задач [12]. Технологии, внедренные на СК «Говорун», позволили развить ИТ-решения, такие как экосистема ML/DL/HPC, предоставляющие возможности не только для решения задач в области машинного и глубокого обучения, но и для удобной организации проведения расчетов и анализа результатов. Примерами таких решений служат разработанная информационно-вычислительная система для совместного проекта с ЛТФ по изучению теоретических моделей джозефсоновских переходов [12] и информационная система для совместного проекта с ЛРБ для обработки, анализа и визуализации данных радиобиологических исследований [13].

Еще одним направлением исследований, в котором задействованы ресурсы СК «Говорун», является создание объединенной масштабируемой научно-исследовательской суперкомпьютерной инфраструктуры на базе Национальной исследовательской компьютерной сети России (НИКС). В настоящее время в эту инфраструктуру, помимо СК «Говорун», входят суперкомпьютеры Межведомственного суперкомпьютерного центра Российской академии наук и Санкт-Петербургского политехнического университета Петра Великого. Созданная инфраструктура позволяет участникам расширять свои локальные вычислительные мощности, обеспечивать доступ к средствам хранения и обработки больших объемов данных, к распределенным хранилищам данных (датахабам), а также использовать мощности друг друга в случаях пиковых нагрузок. Такая инфраструктура востребована в первую очередь для задач мегасайнс проекта NICA. В 2022 году успешно завершен первый совместный эксперимент по использованию объединенной суперкомпьютерной инфраструктуры для задач мегасайнс проекта NICA. Всего было запущено 3000 задач генерации данных методом Монте-Карло и реконструкции событий для эксперимента MPD. В результате все задачи были выполнены успешно. Сгенерировано и реконструировано порядка 3

миллионов событий. Полученные данные перемещены в ОИЯИ для дальнейшей обработки и физического анализа.

Результаты, полученные с использованием ресурсов СК «Говорун» с момента ввода его в эксплуатацию с июля 2018 г. по сентябрь 2022 г., отражены в 204 публикациях пользователей, при этом две из них в журнале Nature Physics.

Таким образом, опыт эксплуатации СК «Говорун» показал востребованность и результативность использования как новейших гиперконвергентных вычислительных архитектур, так и входящей в его состав иерархической системы обработки и хранения данных. В настоящее время ресурсы СК «Говорун» используются научными группами из всех Лабораторий Института в рамках 25 тем Проблемно-тематического плана ОИЯИ. Число пользователей СК «Говорун» составляет 323 человека, из них 262 являются сотрудниками ОИЯИ, а 61 - из стран-участниц. Доступ к ресурсам СК «Говорун» предоставляется только тем пользователям, которые принимают непосредственное участие в реализации ПТП ОИЯИ.

Наиболее важными результатами являются:

1. Впервые в мире создана и внедрена гиперконвергентная архитектура для вычислительных узлов суперкомпьютера. Гиперконвергентность вычислительных узлов позволяет «оркестрировать» вычислительными ресурсами и элементами хранения данных и создавать, используя ПО РСК БазИС, вычислительные системы по требованию задач пользователя. Помимо повышения эффективности решения пользовательских задач разных типов, это свойство позволяет повысить уровень конфиденциальности работы с данными и избежать системных ошибок, возникающих при пересечении ресурсов для различных пользовательских задач.
2. Разработана и внедрена иерархическая система обработки и хранения данных, представляющая собой единую централизованно управляемую систему, имеющую несколько уровней хранения данных: очень горячие данные, горячие данные и теплые данные. Использование этого решения позволило сформулировать и реализовать концепцию работы с Большими данными на СК «Говорун» как реализацию отображения (mapping) основных характеристик больших данных V^3 (Volume – большие объемы данных для обработки и хранения, Velocity - необходимость в высокоскоростной их обработки, Variety – данные различных типов) на программно-аппаратные характеристики суперкомпьютера H^3 (Heterogenity – набор вычислителей разного типа, Hierarchy-многоуровневая организация доступа к данным, Hyperconvegence – динамичная организация систем хранения данных). Внедрение иерархической системы обработки и хранения данных позволяет существенно повысить эффективность работы с большими массивами данных, в том числе для проекта NICA.
3. Гибкая архитектура СК «Говорун» дает возможность не только проводить расчеты, но и использовать суперкомпьютер как научно-исследовательский полигон для выработки программно-аппаратных и ИТ-решений для задач, решаемых в ОИЯИ. Это свойство позволило развернуть полигоны для квантовых вычислений и для обработки экспериментальных данных ЛРБ, включить ресурсы СК «Говорун» в единую гетерогенную среду на основе платформы DIRAC для проекта NICA и задействовать его ресурсы для реализации программы сеансов массового моделирования данных эксперимента MPD. Следует отметить, что некоторые задачи для моделирования данных эксперимента MPD возможно выполнить только на ресурсах СК «Говорун».

Список публикаций:

[1] A. Baginyan, A. Balandin, N. Balashov, A. Dolbilov, A. Gavrish, A. Golunov, N. Gromova, I. Kashunin, V. Korenkov, N. Kutovskiy, V. Mitsyn, I. Pelevanyuk, D. Podgainy, O. Streltsova, T. Strizh, V. Trofimov, A. Vorontsov, N. Voytishin and M. Zuev: “Current Status of the MICC: an Overview” // CEUR Workshop proceedings, 2021, Vol. 3041, pp. 1-8.

[2] Gh. Adam, M. Bashashin, D. Belyakov, M. Kirakosyan, M. Matveev, D. Podgainy, T. Sapozhnikova, O. Streltsova, Sh. Torosyan, M. Vala, L. Valova, A. Vorontsov, T. Zaikina, E. Zemlyanaya and M. Zuev: “IT-ecosystem of the HybriLIT heterogeneous platform for high-performance computing and training of IT-specialists” // CEUR Workshop proceedings, 2018, Vol. 2267, pp. 638-644.

[3] D.V. Podgainy, D.V. Belaykov, A.V. Nechaevsky, O.I. Streltsova, A.V. Vorontsov and M.I. Zuev: “IT Solutions for JINR Tasks on the “GOVORUN” Supercomputer” // CEUR Workshop proceedings, 2021, Vol. 3041, pp. 612-618.

[4] E.A. Druzhinin, A.B. Shmelev, A.A. Moskovsky, V.V. Mironov, A. Semin, “Server Level Liquid Cooling: Do Higher System Temperatures Improve Energy Efficiency?” // Supercomputing frontiers and innovations, 2016, Vol. 3, № 1, pp. 67-73, DOI: 10.14529/jsfi160104

[5] E. Druzhinin, A. Shmelev, A. Moskovsky, Yu. Migal, V. Mironov, A. Semin, “High temperature coolant demonstrated for a computational cluster” // Proc. of 2016 International Conference on High Performance Computing & Simulation (HPCS), DOI: 10.1109/HPCSim.2016.7568418

[6] D. Belyakov, A. Nechaevskiy, I. Pelevanuk, D. Podgainy, A. Stadnik, O. Streltsova, A. Vorontsov, M. Zuev: “Govorun” Supercomputer for JINR Tasks” // CEUR Workshop proceedings, 2020, Vol. 2772, pp. 1-12.

[7] V. Korenkov, A. Dolbilov, V. Mitsyn, I. Kashunin, N. Kutovskiy, D. Podgainy, O. Streltsova, T. Strizh, V. Trofimov, and P. Zrelov: “The JINR distributed computing environment” // EPJ Web of Conferences, 2019, Vol. 214, p. 03009, DOI: https://doi.org/10.1051/epjconf/201921403009

[8] В.В. Кореньков: “Тенденции и перспективы развития распределенных вычислений и аналитики больших данных для поддержки проектов класса мегасайенс” // Ядерная физика, 2020, том 83, № 6, с. 534-538.

[9] D.V. Belyakov, AG. Dolbilov, A.A. Moshkin, I.S. Pelevanyuk, D.V. Podgainy, O.V. Rogachevsky, O.I. Streltsova and M.I. Zuev: “Using the “Govorun” Supercomputer for the NICA Megaproject” // CEUR Workshop proceedings, 2018, Vol. 2507, pp. 316-320.

[10] N. Kutovskiy, V. Mitsyn, A. Moshkin, I. Pelevanyuk, D. Podgayny, O. Rogachevsky, B. Shchinov, V. Trofimov and A. Tsaregorodtsev: “Integration of Distributed Heterogeneous Computing Resources for the MPD Experiment with DIRAC Interware”// Physics of Particles and Nuclei, 2021, Vol. 52 (4), pp. 835-841, DOI:10.1134/S1063779621040419

[11] A.A. Moshkin, I.S. Pelevanyuk, D.V. Podgainy, O.V. Rogachevsky, O.I. Streltsova and M.I. Zuev: “Approaches, services, and monitoring in a distributed heterogeneous computing environment for the MPD experiment” // Russian Supercomputing Days: Proceedings of the International Conference, 2021, pp. 4-11. DOI: https://doi.org/10.29003/m2454.RussianSCDays2021.

[12] Ю.А. Бутенко, М.И. Зуев, М. Чосич, А.В. Нечаевский, Д.В. Подгайный, И.Р. Рахмонов, А.В. Стадник, О.И. Стрельцова. Экосистема ML/DL/HPC платформы HybriLIT (ЛИТ ОИЯИ): новые возможности для прикладных исследований // (направлена в печать), 2022.

[13] I.A. Kolesnikova, AV. Nechaevskiy, D.V. Podgainy, A.V. Stadnik, A.I. Streltsov and O.I. Streltsova: “Information System for Radiobiological Studies” // CEUR Workshop proceedings, 2020, Vol. 2743, pp. 1-6.