

Вызовы и перспективы суперкомпьютерных технологий в России



Святослав СМИРНОВ,
руководитель подразделения «K2 НейроТех»

Долгое время Россия уступала пальму первенства в сфере HPC-технологий западным странам и Китаю. Российский рынок высокопроизводительных вычислений по-прежнему заметно отстает от мирового уровня. Лидерами являются США и Китай, где суперкомпьютеры уже вышли на уровень эксафлопс, тогда как лучшие российские системы лишь приближаются к этим показателям. Причины отставания – недостаточно развитая электронная компонентная база, проблемы с финансированием и конкуренция с общемировыми лидерами, которые фактически монополизировали рынок.

ПОЗИТИВНЫЕ СДВИГИ

Тем не менее, в российском сегменте HPC наметились позитивные сдвиги. Крупные корпорации и университеты активнее сотрудничают с государственными

Стремительное развитие цифровой экономики ставит перед странами новые задачи в сфере информационных технологий. Если раньше мощные компьютеры использовались преимущественно в науке, исследованиях и промышленности, то теперь высокопроизводительные вычисления (HPC) становятся ключевым инструментом для бизнеса и госуправления. Быстрый рост сегмента заставляет многие государства пересматривать подходы к стратегии формирования инфраструктуры. Для России, пока отстающей в этой сфере, ситуация осложняется целым рядом факторов. Однако наметились сдвиги к улучшению.

органами власти, развивая совместные проекты по ускоренному развитию суперкомпьютерных технологий. Реализован ряд интересных инициатив, таких как суперкомпьютер МГУ «Ломоносов-2», суперкомпьютер Christofari Neo от Сбера и суперкомпьютер «Червоненкис» от Яндекса, которые представляют собой крупнейшие вычислительные системы в стране. Такие проекты подтверждают растущее значение HPC для технологической независимости и конкурентоспособности России на международной арене.

Один из важнейших шагов к развитию национального HPC-потенциала – создание благоприятной законодательной и регуляторной среды. Власти приняли меры по снижению налоговой нагрузки для компаний, занимающихся разработкой и производством компонентов для суперкомпьютеров. Такая политика должна стимулировать привлечение инвестиций и расширять круг игроков на российском рынке HPC.

Еще один аспект – разработка образовательных программ и подготовка специалистов с нужными компетенциями. В условиях стремительного развития HPC-технологий крайне важно, чтобы

будущие инженеры обладали глубокими познаниями в математике, физике и информатике, необходимыми для работы с такими сложными системами.

Параллельно предстоит решить задачи в сфере защиты интеллектуальной собственности и стандартов безопасности. Сотрудничество с иностранными партнерами может дать значительные преимущества, но при этом следует сохранять контроль над внутренними проектами и предотвращать утечку конфиденциальной информации. Для этого необходимы усиление правового регулирования, контроль за экспортом чувствительных технологий и тщательная проверка используемых комплектующих.

В целом перспективы развития российского рынка HPC выглядят обнадеживающе. У страны есть значительный научный потенциал и государственный интерес к этой сфере. Тем не менее, достижение лидерских позиций потребует существенных ресурсов и усилий как частного сектора, так и органов власти. Лишь объединяя силы и целенаправленно развивая инфраструктуру, Россия сможет претендовать на заметную роль в глобальной гонке высокопроизводительных вычислений.

Препятствия в сегменте HPC

Из топ-500 лучших суперкомпьютеров мира Россия представлена всего несколькими позициями, причем большинство из них относится к низкоранговым устройствам, предназначенным для специализированных целей.

Одним из ключевых препятствий развития российского HPC-рынка остается слабая национальная элементная база. Отечественные разработчики не способны производить конкурентоспособные процессоры и ускорители, сопоставимые по характеристикам с продукцией Intel, AMD или Nvidia. Даже простейшие комплектующие, такие как модули оперативной памяти и интерфейсы связи, зачастую импортируются из-за рубежа, что снижает независимость российской индустрии и увеличивает стоимость готовой продукции.

Однако на рынке представлены аналоги, в том числе китайские и российские. По производительности отечественные решения находятся пока на уровне лидеров пяти–семилетней давности и подходят для задач инференса или обучения небольших моделей.

Тем не менее, позитивные изменения есть. Крупные научные центры – МГУ им. М.В. Ломоносова, Курчатовский институт и Новосибирский Академгородок – запустили пилотные проекты по строительству небольших суперкомпьютерных кластеров для научных и коммерческих задач. Кроме того, частные компании предлагают услуги аренды вычислительных мощностей для бизнеса.

Растет число малых и средних предприятий, создающих собственное ПО для HPC-систем. Некоторые из них специализируются на создании отраслевого программного обеспечения для нефтяной и газовой промышленности, машиностроения и здравоохранения. Хотя общий уровень проникновения HPC-технологий в экономику России пока невысок, заметен существенный рост в этой области, дающий надежду на сокращение отставания от мировых лидеров.

Результативные инициативы

За последние годы отечественные разработчики добились значимых результатов в создании конкурентоспособных решений. Одна из наиболее успешных инициатив – строительство нового поколения суперкомпьютеров с высокой производительностью и энергоэффективностью. Среди заметных разработок:

Российский рынок HPC делится на несколько сегментов, в каждом представлены специализированные компании и учреждения. Многоложность структуры обусловлена необходимостью сотрудничества различных секторов экономики для достижения наилучших результатов. Среди ключевых участников – разработчики ПО и оборудования, системные интеграторы, научные институты и вузы.

В целом перспективы развития российского рынка HPC выглядят обнадеживающе. У страны есть значительный научный потенциал и государственный интерес к этой сфере.

- специализированные GPU-платформы. Например, сервер YADRO G4208P G3. Продукт позиционируется как высокоплотный GPU-сервер, позволяющий устанавливать до восьми ускорителей стандарта PCIe 5.0 x16. Благодаря высокой плотности и модульности дизайна этот сервер – привлекательный вариант для корпоративных клиентов, заинтересованных в обучении нейросетей и обработке больших массивов данных;
- CPU-ориентированные решения, например Эльбрус-8СВ, Эльбрус-16С и Байкал-М8. Эти технологии обеспечивают исключительную вычислительную мощность для традиционных HPC-задач, включая сложное моделирование и научные исследования;
- комплексные решения. Например, ПАК-HPC от компании «К2 НейроТех» объединяет аппаратную платформу с моделями сервисов, предоставляя клиентам полный набор услуг по использованию вычислительных ресурсов. Особенность проекта – готовое решение, включающее в себя единое окно техподдержки, портал самообслуживания, биллинг и все необходимые инструменты и сервисы.

Отраслевые тенденции

Глобальные изменения в мировой HPC-индустрии вызваны появлением новых видов оборудования и программного обеспечения. Можно выделить несколько ключевых трендов, оказывающих влияние на общее развитие отрасли.

Первый – активная модернизация и масштабирование HPC-кластеров. Созданные 10–15 лет назад суперкомпьютеры уже не отвечают текущим требованиям: устаревшее «железо» сложно поддерживать, а многие производители покинули рынок. Это заставляет компании не просто обновлять оборудование, а фактически перестраивать архитектуру HPC-платформ с нуля. Переход на новые компоненты, актуальное ПО и адаптация к современным нагрузкам позволяют устранить узкие места и обеспечить гибкость для будущего роста.

Одно из ключевых направлений в развитии HPC – переход на отечественные компоненты. Это стратегическое решение для многих отраслей, где сбои недопустимы. Вычислительная платформа должна быть максимально адаптирована к локальным реалиям.

Результат таких изменений очевиден: современные кластеры становятся более гибкими, удобными в обслуживании и устойчивыми к внешним рискам. Компании, решившиеся на обновление, не только «догоняют» растущие требования к вычислениям, но и формируют прочную технологическую основу на будущее.

Вторая тенденция – переход к таким гибридным архитектурам, как CPU, GPU и специализированные ускорители. Современные HPC-задачи требуют высокой мощности и точного распределения нагрузки между различными типами оборудования. Для разных

El Capitan построен на базе APU AMD Instinct MI300A, объединяющим в одном корпусе процессорные ядра Zen 4 (CPU) и графические блоки CDNA 3 (GPU), а также общую память типа HBM3. Унифицированная архитектура позволяет уменьшить задержки и ускорить передачу данных между компонентами. Каждый MI300A содержит 24 ядра Zen 4 и 228 вычислительных блоков GPU. В каждом узле El Capitan установлено несколько таких чипов. Заявленная пиковая производительность системы составляет 2,79 эксафлопса, а по последнему рейтингу топ-500 ее измеренная производительность – 1,742 эксафлопса.

В настоящее время El Capitan – самый мощный суперкомпьютер.

Другой пример – китайский Tianhe-3 (Xingyi), разработанный Национальным университетом оборонных технологий (NUDT). Хотя официальные данные о нем частично засекречены, из материалов издания The Next Platform известно, что в его основе – процессоры MT-3000 с гетерогенной архитектурой. Каждый чип включает 16 ядер общего назначения, 96 управляющих ядер и 1536 ядер-ускорителей. По оценкам экспертов рынка, Tianhe-3 способен достигать производительности до 1,57 эксафлопса.

Перспективные компоненты и решения

Помимо проверенных решений развивается направление новых типов ускорителей – квантовых и нейроморфных чипов – как перспективных компонентов гибридных систем.

Например, квантовые процесоры IBM серии Eagle и Osprey обеспечивают выполнение алгоритмов с повышенной устойчивостью к ошибкам, что открывает дополнительные возможности для решения задач факторизации и оптимизации, недоступных классическим HPC-платформам.

Нейроморфные процесоры Intel Loihi и исследовательские разработки IBM TrueNorth реализуют архитектуры, основанные на спайковом нейронном моделировании, что обеспечивает высокую энергоэффективность и ускорения при глубинном обучении и обработке потоковых данных. Пока подобные направления на стадии экспериментов, и они формируют задел для решения задач, связанных с генеративным ИИ и моделированием сложных систем, в будущем это может значительно расширить функционал HPC-систем.

Переход к гибридным архитектурам – не просто модернизация, а переосмысление подхода к распределению вычислительных задач, что позволяет создавать более гибкие, энергоэффективные и адаптивные вычислительные платформы.

При этом собственная HPC-инфраструктура не всегда оправдана, особенно если задачи возникают нерегулярно и носят краткосрочный характер. В таких случаях все большую популярность набирает модель HPC-as-a-Service, когда вычислительные ресурсы предоставляются как сервис. Примером решения, которое позволяет компаниям внедрять эту модель внутри инфраструктуры, может служить ПАК-HPC от «К2 НейроТех». К его преимуществам относятся централизованное управление, гибкое выделение ресурсов и предоставление готовых инструментов для параллельных вычислений.

К модели HPC-as-a-Service прибегают организации, которым важно централизованно управлять доступом к вычислительным ресурсам, поддерживать стандартизированные рабочие окружения и обеспечивать непрерывность

типов нагрузок нужно специализированное оборудование. Там, где критична высокая последовательная производительность, доминируют CPU, а там, где требуется массовое параллельное исполнение, – GPU. Такой подход позволяет «разложить» вычислительные задачи по оптимальным типам узлов, обеспечивая точную настройку кластера под численное моделирование, расчеты в химии и физике, визуализацию или задачи материаловедения. Международная практика и отчеты технологических трендов 2025 г., в том числе CIQ Empowers Researchers to Innovate Faster with Fuzzball, демонстрируют, что уже сегодня гибридные архитектуры активно применяются для повышения эффективности вычислений.

Примеры суперкомпьютеров

Запущенный в 2024 г. в Национальной лаборатории Лоуренса Ливермора (США) суперкомпьютер

процессов. Это может быть как промышленное предприятие с периодическими нагрузками, так и вуз, научный центр или ИТ-департамент, где расчеты и анализ данных требуются по мере возникновения задач.

При интеграции локальных кластеров с облачными сервисами возможно оперативно масштабировать вычислительную базу. Это позволяет адаптироваться к нестабильности нагрузки, оптимизировать затраты, выбирая ресурсы, требуемые для решения конкретной задачи. Такой подход исключает необходимость крупных капитальных вложений в собственные data-центры, инфраструктура становится более управляемой и предсказуемой в эксплуатации.

Контейнеризация в HPC – спасение для тех, кто устал от не-гибкости монолитных решений. В условиях, когда множество пользователей одновременно работают с разными приложениями, библиотеками и окружениями, традиционный подход приводит к конфликтам версий, затяжным развертываниям и сложностям в поддержке. Контейнеры решают эти задачи за счет изоляции среды, предсказуемости выполнения и быстрого тиражирования решений.

В HPC-кластерах «К2 НейроТех», например, контейнерные технологии интегрированы в платформу управления, построенную по принципам IaaS, PaaS и SaaS. При этом в IaaS-среде, наряду с классическими параллельными расчетами на полном кластере, разработчику часто требуется не весь кластер, а отдельная виртуальная машина с GPU, обеспечивающая гибкость и экономию ресурсов. На уровне PaaS контейнеры используются для развертывания готовых вычислительных окружений с планировщиком, авторизацией, мониторингом и управлением через API. Это позволяет оперативно подстраиваться под конкретную задачу, например численное моделирование, обработку больших данных, машинное обучение и визуализацию результатов. На уровне SaaS

запускаются специализированные прикладные инструменты – среды визуализации (например ParaView, VisIt), отладки (GDB, NVIDIA Nsight), аналитики (Apache Spark, Jupyter Notebook) и расчета (MATLAB, ANSYS). Каждая из них функционирует в собственном изолированном контейнере.

отечественных процессоров. Современные суперкомпьютеры построены на базе процессоров Xeon от Intel или EPYC от AMD, которые превосходят российские аналоги по производительности. Без качественных процессоров эффективное производство HPC невозможно;

Развивается направление новых типов ускорителей – квантовых и нейроморфных чипов – как перспективных компонентов гибридных систем.

Такой подход оптимизирует эксплуатационные расходы, упрощает управление нагрузками и позволяет быстро адаптировать инфраструктуру под изменяющиеся требования. HPC давно перестали быть нишевой технологией для научных институтов: сегодня это полноценный инструмент для производственных процессов, проектирования, разработки новых материалов и анализа данных.

Особый интерес вызывает область экспериментальных исследований, включающая квантовые и нейроморфные процессоры. Данные технологии находятся на ранних стадиях развития, но уже очевиден их большой потенциал для решения задач, недоступных классическим вычислительным системам.

Выборочная локализация и комбинирование

Российские специалисты демонстрируют определенную степень компетентности в области HPC, но перед ними стоят серьезные вызовы:

- процессорная техника. Основное препятствие для выхода России на мировой рынок HPC – отсутствие конкурентных

- специализированные компоненты. Не хватает отечественных ускорителей для задач искусственного интеллекта и машинного обучения. Большие мощности, необходимые для обработки больших объемов данных, оказываются недоступны, поскольку отечественные разработки не соответствуют мировым стандартам;

- интерконнекты и периферия. Дефицит эффективных каналов передачи данных и интерфейсных устройств. Большая часть высокоскоростных сетей Ethernet и InfiniBand производится за рубежом, приобретение их сопряжено с техническими трудностями и политическими ограничениями;

- операционная среда и ПО. Современные операционные системы и компиляторы для HPC тоже разрабатываются за рубежом. Российским компаниям сложно создавать подходящие аналоги и поддерживать их качество на должном уровне.

Политика импортозамещения была принята Правительством России в качестве основного курса развития ИТ-отрасли. Ее цель – максимальное использование внутренних ресурсов и минимизация зависимости от внешнего рынка. Изначально

предполагалось, что полная замена иностранной техники приведет к быстрому восстановлению утраченных позиций. Однако оказалось, что задача значительно сложнее.

Основная причина неудач кроется в медленной адаптации российской промышленности к новым условиям. Многие предприятия

В России развитие ИИ-технологий обеспечивается преимущественно усилиями университетов и корпораций. Ведущие участники рынка – крупнейшие компании, работающие над созданием собственных систем распознавания речи, рекомендаций товаров и анализа социальных медиа. Ведущие вузы – ВШЭ, МФТИ

климата и медицины, нуждаются в сверхбыстрых вычислениях для моделирования природных явлений и создания предсказательных моделей. Университеты и НИИ регулярно обращаются к услугам НРС-центров для проведения научных экспериментов и подготовки публикаций;

- банковский сектор. Банки и инвестиционные компании используют НРС для высокочастотного трейдинга, кредитного скоринга и управления рисками. Вычислительная мощность необходима для точного прогнозирования рыночных колебаний и снижения возможных убытков;
- телекоммуникации. Операторы мобильной связи и интернет-провайдеры полагаются на НРС при анализе трафика, планировании сети и предотвращении перегрузок. Чем выше нагрузка на сеть, тем важнее точные прогнозы и принятие правильных управлений решений.

Таким образом, российский рынок НРС характеризуется широким разнообразием запросов и потенциальных пользователей. Отсутствие достаточной инфраструктуры замедляет развитие отрасли. Отечественный сегмент НРС сейчас на сложном этапе, полном вызовов и возможностей. Корректировка политики импортозамещения позволила сохранить движение вперед.

Однако впереди немало преград, в частности, низкий уровень производства комплектующих, необходимость инвестиций в обучение молодых специалистов. Преодоление этих препятствий позволит России обрести прочное положение на международном рынке НРС и усилить присутствие в глобальной экономике. Необходимо и дальше прилагать усилия с целью совершенствования законодательства, защиты прав интеллектуальной собственности и обеспечения условий для процветания малого и среднего бизнеса. Совместными усилиями удастся преодолеть кризис и заложить основы устойчивого развития отрасли на долгие годы. ■

Один из ключевых драйверов развития НРС-рынка – повсеместный интерес к искусственному интеллекту.

оказались неспособны перестроиться и освоить производство необходимого оборудования, что заставило Правительство пересмотреть стратегию импортозамещения. Сегодня акцент смешен с полной замены иностранной техники на выборочную локализацию и комбинирование российских разработок с зарубежными решениями.

Более гибкий и реалистичный подход позволяет сохранять необходимую долю зарубежного оборудования, но при этом отдавать приоритет собственным технологиям, что должно способствовать увеличению доли российских решений на рынке и повышению конкурентоспособности отечественной НРС-отрасли.

ИИ как драйвер развития

Один из ключевых драйверов развития НРС-рынка – повсеместный интерес к искусственному интеллекту. Применение ИИ-технологий расширилось настолько, что сегодня ни одна крупная организация не обходится без их использования. Высокоразвитые НРС-системы позволяют ускорить выполнение задач машинного обучения и анализа больших данных, создавая огромные возможности для бизнеса и науки.

и СПбГУ – готовят специалистов по искусственному интеллекту и участвуют в международных проектах, посвященных применению этой технологии.

Особенность российской специфики – наличие огромного массива исторических данных, накопленных во времена СССР. Эти данные представляют большую ценность для исследователей и ученых, желающих применять ИИ-методы для решения текущих задач. Благодаря этому НРС-индустрия в России получила мощный стимул, вызванный спросом на высокопроизводительные вычисления для анализа информации.

Основные потребители НРС-вычислений в России – представители следующих ключевых секторов экономики:

- автомобильная, авиационно-космическая и энергетическая отрасли промышленности, которые традиционно испытывают потребность в расчетах на высокопроизводительных системах. Моделирование поведения материалов, проектирование конструкций и оценка воздействия окружающей среды требуют значительных вычислительных мощностей;
- наука и образование. Ученые, занятые исследованием космоса,