УДК 004.72, 004.75, 004.382.2 DOI: 10.14529/cmse200102

ОБ ИСПОЛЬЗОВАНИИ ФЕДЕРАЛЬНОЙ НАУЧНОЙ ТЕЛЕКОММУНИКАЦИОННОЙ ИНФРАСТРУКТУРЫ ДЛЯ СУПЕРКОМПЬЮТЕРНЫХ ВЫЧИСЛЕНИЙ*

© 2020 Г.И. Савин, Б.М. Шабанов, А.В. Баранов, А.П. Овсянников, А.А. Гончар

Межведомственный суперкомпьютерный центр РАН — филиал ФГУ ФНЦ НИИСИ РАН (119334 Москва, Ленинский пр., д. 32а)

E-mail: savin@jscc.ru, shabanov@jscc.ru, antbar@mail.ru, ovsiannikov@jscc.ru, andrey.gonchar@jscc.ru

Поступила в редакцию: 09.01.2020

Статья посвящена перспективам развития научной телекоммуникационной инфраструктуры на базе национальной исследовательской компьютерной сети нового поколения (НИКС), образованной путем интеграции ведомственных научно-образовательных сетей RUNNet и RASNet. Показаны возможности новой сети для объединения и организации взаимодействия суперкомпьютерных ресурсов и обеспечения безбарьерного доступа к ним. На основе обобщенного мирового опыта показано, что суперкомпьютерные инфраструктуры предъявляют специальные требования к телекоммуникационной сети по передаче данных и наличию ряда дополнительных сервисов. Эти требования выходят далеко за рамки услуг коммерческих операторов связи и, как правило, могут быть удовлетворены только объединенными усилиями национальных научно-образовательных сетей.

Рассмотрены ключевые элементы федеральной телекоммуникационной инфраструктуры, необходимые для объединения высокопроизводительных вычислительных ресурсов: высокопроизводительные каналы связи с заданным качеством обслуживания, их автоматическое выделение по требованию и по расписанию, доверенная сетевая среда, федеративная аутентификация и авторизация, обеспечение надежности и безопасности, сквозной мониторинг пути передачи данных между конечными пользователями. На основе анализа жизненного цикла суперкомпьютерного задания, мигрирующего в сети суперкомпьютерных центров коллективного пользования (СКЦ), сформулированы требования к телекоммуникационной инфраструктуре НИКС и сервисам на ее основе со стороны распределенной сети СКЦ.

Ключевые слова: национальная сеть науки и образования, суперкомпьютерный центр, центр коллективного пользования, распределенные вычисления, телекоммуникационная инфраструктура.

ОБРАЗЕЦ ЦИТИРОВАНИЯ

Савин Г.И., Шабанов Б.М., Баранов А.В., Овсянников А.П., Гончар А.А. Об использовании федеральной научной телекоммуникационной инфраструктуры для суперкомпьютерных вычислений // Вестник ЮУрГУ. Серия: Вычислительная математика и информатика. 2020. Т. 9, № 1. С. 20–35. DOI: 10.14529/cmse200102.

Введение

Наличие собственной надежной и высокопроизводительной телекоммуникационной инфраструктуры для науки и образования, обеспечивающей доступ к высокопроизводительным вычислительным ресурсам, суперкомпьютерным центрам, их сетевое взаимодействие, является сегодня необходимым условием для выхода на мировой уровень достижений в самых различных областях науки и технологий.

*Статья рекомендована к публикации программным комитетом международной конференции «Суперкомпьютерные дни в России — 2019».

Как отмечалось в [1], инфраструктура суперкомпьютерных центров включает не только вычислительную технику различной архитектуры, но и средства визуализации, информационные ресурсы и систему телекоммуникаций, причем приоритетное значение придается развитию систем обеспечения удаленного доступа со сбалансированной иерархией сетей различной пропускной способности.

Удаленный доступ к ресурсам суперкомпьютерного центра коллективного пользования (СКЦ) осуществляется как по выделенным сетям с высокой пропускной способностью (десятки гигабит в секунду), так и через публичные сети. Первые позволяют передавать значительные объемы данных между системами хранения данных (СХД) организаций-пользователей и СХД СКЦ, вторые — запускать задания, используя данные, заранее размещенные в СХД СКЦ, и там же сохранять результаты вычислений.

Мировой опыт в области создания и развития научно-образовательных телекоммуникационных сетей свидетельствует об особой роли в организации информационного обмена СКЦ национальных научно-образовательных сетей (или национальных сетей науки и образования — National Research and Education Network, NREN).

Национальные научно-образовательные сети эволюционируют, откликаясь на возрастающие потребности науки и образования, оперативно реагируя на лавинообразный рост объемов генерируемых и обрабатываемых данных, требующих передачи по сети, на интенсивное развитие средств телекоммуникаций, появление и широкое внедрение перспективных информационно-коммуникационных технологий и сервисов. Их цели, ключевые задачи и основные функции непрерывно уточняются и трансформируются.

Статья организована следующим образом. В разделе 1 обобщен мировой опыт использования национальных сетей науки и образования для суперкомпьютерной инфраструктуры, в том числе рассмотрены крупнейшие проекты PRACE и XSEDE. Раздел 2 посвящен процессу интеграции ведомственных научно-образовательных сетей RASNet и RUNNet в национальную сеть науки и образования. В разделе 3 рассмотрены вопросы создания прикладной цифровой платформы, объединяющей в единую сеть вычислительные ресурсы территориально распределенных суперкомпьютерных центров коллективного пользования. В разделе 4 определены требования, предъявляемые к национальной сети науки и образования со стороны создаваемой прикладной цифровой платформы. В заключении кратко обобщены сформулированные требования и намечены направления дальнейших исследований.

1. Мировой опыт использования национальных сетей науки и образования для суперкомпьютерной инфраструктуры

К настоящему моменту национальные сети науки и образования существуют в большинстве стран мира, функционируют в качестве неотъемлемой части национальной ИКТ-инфраструктуры, как правило, координируются государственными органами управления наукой и образованием, представляют страну в международных проектах, при реализации которых интенсивно используются современные средства телекоммуникаций, развитые сетевые технологии и прикладные сервисы.

Примерами таких сетей являются DFN (Германия) [2], CANARIE (Канада) [3], Internet2 (США) [4], SURFnet (Нидерланды) [5], AARnet (Австралия) [6], CERNET (Китай) [7].

Для эффективной реализации международных проектов и взаимной кооперации создан ряд наднациональных объединений — консорциумов научно-образовательных сетей:

NORDUnet (Скандинавские страны) [8], GÉANT (Европа) [9], Asi@Connect/TEIN [10] и APAN (Азия) [11], RedClara (Южная Америка) [12], AfricaConnect [13].

Национальные сети науки и образования и их сообщества всегда использовались в качестве транспортной инфраструктуры для проектов объединения суперкомпьютерных ресурсов, например, проектов TeraGrid [14] и XSEDE [15] в США, DEISA [16] и PRACE [17] в Европе, NAREGI [18] в Японии.

Для таких проектов национальные сети науки и образования и их международные объединения развивают сервисы передачи данных со специальными требованями по качеству обслуживания: пропускной способности, уровню задержек и потерь пакетов, надежности работы и резервированию каналов связи. Ведутся разработки средств оперативного (в режиме онлайн) управления сетевыми ресурсами научных телекоммуникационных сетей: выделения и настройки каналов связи между научными организациями по требованию и расписанию.

Проект TeraGrid (2001–2011) [14] использовал выделенную волоконно-оптическую магистральную сеть 40 Гбит/с между Национальной лабораторией в Аргонне (Argonne National Laboratory), Калифорнийским технологическим институтом (California Institute of Technology), Национальным центром суперкомьютерных приложений (National Center for Supercomputing Applications) и Суперкомпьютерным центром в Сан-Диего (San Diego Supercomputer Center). Пользователи получали доступ к ресурсам проекта через национальные исследовательские сети: Internet2, Abilene backbone и National LambdaRail.

В проекте DEISA (2002–2011) [16] организации, предоставлявшие суперкомпьютерные ресурсы, подключались на скорости 10 Гбит/с к сети, которая была организована совместно общеевропейской магистральной научно-образовательной сетью GÉANT и национальными сетями науки и образования стран, участвовавших в проекте DEISA.

Для проекта PRACE — Partnership for Advanced Computing in Europe (2010–наст. время) — сетевой консорциум GÉANT построил новую высокоскоростную сеть на основе сервиса Многодоменной виртуальной частной сети (Multidomain Virtual Private Network, MD-VPN) [19].

Особенностью частной сети MD-VPN является организация наложенных сетей L3VPN (IPv4/IPv6), каналов L2VPN и многоточечных сетей L2VPN через несколько сетевых провайдеров (доменов) [20]. Конечным пользователям сервис предоставляется через так называемую точку демаркации услуг (Service Demarcation Point — SDP) на границе национальной сети науки и образования или даже региональной сети. На практике способ предоставления услуги конечным пользователям зависит от национальной сети науки и образования, обычно это L3VPN в форме IP-пакетов на третьем (сетевом) уровне и VLAN на порту или в пакете 802.1q для коммутируемых операторских сетей на основе протокола Ethernet или двухточечный L2VPN/многоточечный VPLS для сетей MPLS. В результате конечные пользователи могут работать в сетях IPv4/IPv6 или Ethernet так, как если бы их сети были связаны напрямую (промежуточные сети прозрачны для конечных пользователей).

Таким образом, суперкомпьютерные центры в разных странах могут работать так, как если бы они находились географически в одном месте; при этом их сети имеют равные уровни безопасности. Это позволяет избежать использования защитных экранов с контролем пакетов по их содержимому (deep packet inspection, DPI), как это потребовалось бы при использовании обычного протокола IP, увеличивает производительность сети и снижает за-

держки. Этот фактор очень важен для распределенных Grid-инфраструктур, облачных и высокопроизводительных вычислений.

Отметим, что служба MD-VPN также обеспечивает более высокий уровень конфиденциальности по сравнению с обычным VPN, так как потоки данных клиента MD-VPN изолированы от любого другого трафика: стандартного IP-трафика и трафика других клиентов MD-VPN.

Наиболее сложные требования к научно-образовательной сети предъявляет проект Extreme Science and Engineering Discovery Environment — XSEDE (2011–наст. время) [21]. В части обеспечения передачи данных требования включают:

- возможность передачи больших объемов данных, включая большие файлы и большое число файлов;
- информирование в реальном времени о состоянии сети и его изменениях в машиночитаемых форматах для взаимодействия с автоматизированными службами XSEDE;
- наличие механизмов обеспечения зарезервированной пропускной способности сети для отдельных пользователей или групп пользователей;
- наличие интегрированных со службами динамического распределения ресурсов сети механизмов управления для временного повышения скорости передачи данных для отдельных пользователей или групп пользователей;
- реализацию и поддержку служб и распределенных компонентов XSEDE и интерфейсов к ним непосредственно в телекоммуникационной сети.

В части сетевой безопасности требования включают:

- наличие общей системы обнаружения вторжений, которая сопоставляет информацию с различных датчиков на сетевом и системном уровне и способна отправлять уведомления в режиме реального времени сотрудникам службы безопасности XSEDE;
- наличие системы сканирования уязвимостей.

Кроме того, XSEDE предъявляет ряд требований по сетевой связности: подключение к телекоммуникационным сетям, поддерживаемым Национальным научным фондом США и правительственными агентствами DOE, DOD, NASA, NIH, зарубежным национальным научно-образовательными сетям, и тесное сотрудничество с ними в области развития сетевых технологий и сервисов.

Таким образом, суперкомпьютерные инфраструктуры предъявляют специальные требования к телекоммуникационной сети по передаче данных и наличию ряда дополнительных сервисов. Эти требования выходят далеко за рамки услуг коммерческих операторов связи и, как правило, могут быть удовлетворены только объединенными усилиями национальных научно-образовательных сетей.

2. Создание национальной компьютерной сети нового поколения на основе интеграции ведомственных научно-образовательных сетей

С начала 1990-х гг. в Российской Федерации при целевой поддержке федеральных органов управления образованием и наукой и иных государственных органов создавались и на протяжении долгих лет успешно эксплуатировались в интересах научно-образовательного сообщества отраслевые информационно-телекоммуникационные сети.

Российское научно-образовательное телекоммуникационное пространство исторически строилось в виде ряда взаимодействующих относительно независимых сетей, в значи-

тельной степени базирующихся на собственной канальной структуре и собственных же международных каналах. Некоторые из таких сетей изначально задумывались и проектировались как сети общего назначения, а другие — как специализированные предметно-ориентированные сети.

На конец 2018 года наиболее значимыми отраслевыми сетями сферы образования и науки являлись:

– Университетская сеть федерального масштаба RUNNet объединяет около 120 организаций высшего образования и науки из всех федеральных округов. Среди них — 70 ведущих университетов, в числе которых МГУ им М.В. Ломоносова, 12 университетов Программы 5/100 и 19 национальных исследовательских университетов.
– Система региональных сетей отделений РАН — RASNet. Наибольшее представительство пользователей RASNet сосредоточено в Московском регионе, где пользователями сети являются около 100 научных организаций.

В связи с разделением в 2018 году Минобрнауки России на два ведомства оператору сети RASNet (Межведомственный суперкомпьютерный центр РАН — МСЦ РАН) и оператору RUNNet (Центр реализации государственной образовательной политики и информационных технологий) было предложено инициировать мероприятия по объединению сетей на базе МСЦ РАН для создания национальной исследовательской компьютерной сети нового поколения (НИКС), которая будет играть роль национальной научно-образовательной сети России на международной арене. Концепция создания НИКС была одобрена на Заседании Совета Министерства науки и высшего образования Российской Федерации по информационно-телекоммуникационной инфраструктуре, информационной безопасности и суперкомпьютерным технологиям 24 апреля 2019 года.

МСЦ РАН обладает многолетним опытом управления глобальной ИКТ-инфраструктурой специализированного назначения, ее поддержки и эксплуатации, который включает управление сетью RASNet на федеральном уровне, обеспечение взаимодействия российских и зарубежных сетей науки и образования, представление страны в международных научных проектах, связанных с развитием научной телекоммуникационной и суперкомпьютерной инфраструктуры (GÉANT, DEISA, PRACE).

Передачу проекта RUNNet в МСЦ РАН и интеграцию сетей RASNet и RUNNet в национальную сеть науки и образования предполагается завершить до конца 2019 года.

Создание и развитие национальной исследовательской компьютерной сети нового поколения должно обеспечить функциональные возможности организациям науки и образования в следующих направлениях:

- сетевая связность и управление сетью:
  – удаленный доступ с обеспечением гарантированного качества сервиса и надежности для научно-образовательных организаций и исследовательских организаций промышленности к установкам класса «мегасайенс», уникальным научным установкам, центрам коллективного пользования научным оборудованием, суперкомпьютерным центрам, научным данным и результатам исследований (в т.ч. сверхбольшим объемам данных), ресурсам организаций науки и образования и сервисам с использованием собственной инфраструктуры и инфраструктуры мировых научно-образовательных сетей;
  – организация постоянной или по требованию/расписанию передачи данных со специальными условиями, в т.ч. выделение каналов связи и организация наложен-

ных сетей с заданным качеством обслуживания (пропускная способность, задержки, потери, надежность, резервирование), конфиденциальности и безопасности для совместных научных проектов, а также корпоративных сетей научно-образовательных организаций, объединяющих отделения и филиалы;
  - сквозной мониторинг пути передачи данных и виртуальных частных сетей, включая предоставление информации в машиночитаемых форматах для взаимодействия с автоматизированными службами;
  - интерфейс управления телекоммуникационными ресурсами и сервисами (заказ, выделение, мониторинг, учет для прикладных цифровых платформ, разрабатываемых организациями науки и высшего образования);
- безбарьерный доступ и безопасность:
  - обеспечение конфиденциальности передаваемых и обрабатываемых научных данных, организация доверенной среды передачи, обработки и хранения информации;
  - федеративная аутентификация и авторизация для безбарьерного доступа к российским и зарубежным научным ресурсам и данным;
  - защита персональных данных;
  - мобильность пользователей и безопасный доступ к научным данным и ресурсам из публичных сетей;
  - предоставление интерфейсов и инструментов федеративного доступа и безопасности для прикладных цифровых платформ, создаваемых и развиваемых организациями науки и образования.

Создаваемая национальная исследовательская компьютерная сеть нового поколения должна интегрироваться в мировое сообщество научно-образовательных сетей в роли российской национальной сети науки и образования, что будет способствовать переносу и распространению в России опыта зарубежных научно-образовательных сетей в развитии телекоммуникационных сервисов и инструментов для поддержки суперкомпьютерных инфраструктур.

3. Проект создания распределенной сети суперкомпьютерных центров коллективного пользования

В настоящее время в МСЦ РАН ведутся работы по созданию прикладной цифровой платформы (ПЦП), объединяющей в единую сеть вычислительные ресурсы территориально распределенных суперкомпьютерных центров коллективного пользования [22] в интересах организаций науки, высшего образования и инновационной деятельности Российской Федерации. Целями подобного объединения являются оптимизация распределения вычислительной нагрузки, упрощение и повышение качества доступа пользователей к высокопроизводительным вычислительным ресурсам, консолидация функций мониторинга и управления распределенными суперкомпьютерными ресурсами.

Для создания платформы необходимы исследования и разработка новых методов и алгоритмов, касающихся всех аспектов функционирования распределенной сети СКЦ: организации единого доступа к пулу объединенных ресурсов, распределения пользовательских заданий по вычислительным ресурсам, мониторинга и прогнозирования состояния вычислительной среды, организации централизованного управления данными, моделирования и анализа работы всей системы.

Создание ПЦП требует разработки и создания:

- децентрализованной автоматизированной системы управления заданиями и ресурсами, которая будет поддерживать глобальную очередь пользовательских заданий и обеспечивать за счет этого оперативное перераспределение вычислительной нагрузки в сети СКЦ;
- единой системы мониторинга, которая позволит оперативно получать информацию о текущих состоянии и загруженности суперкомпьютерных ресурсов распределенной сети;
- единой системы доступа на основе удостоверяющей федерации суперкомпьютерных центров, которая обеспечит пользователей технологией единого входа (Single Sign-On, SSO) на суперкомпьютерные ресурсы распределенной сети, причем авторизоваться в сети можно будет с использованием только личной учетной записи в своей организации;
- общей распределенной системы хранения данных, обеспечивающей единое файловое пространство для всех СКЦ сети.

Создание системы управления сетью СКЦ является одной из ключевых задач проекта. Система управления имеет децентрализованный характер и базируется на схеме асинхронного взаимодействия коллектива равноправных диспетчеров [23]. Взаимодействие диспетчеров осуществляется через единую информационную подсистему, в которой формируется и хранится глобальная очередь пользовательских заданий [24]. Основой информационной подсистемы является распределенная документо-ориентированная СУБД Elasticsearch, причем такой подход позволяет сочетать достоинства централизованной (логическая простота взаимодействия) и децентрализованной (отказоустойчивость и надежность хранения) схем управления.

Несмотря на высокую востребованность отечественных суперкомпьютерных систем, их средняя загрузка часто не превышает 70–80%. При утилизации ресурсов 90% и выше (как, например, в МСЦ РАН или НИВЦ МГУ им. М.В. Ломоносова), загруженность того или иного СКЦ будет определяться средним временем нахождения задания в очереди. Предлагаемая асинхронная децентрализованная схема управления позволит через глобальную очередь организовать оперативное перераспределение заданий в сети СКЦ из более загруженного центра в менее загруженный. За счет этого удастся сократить время пребывания задания в очереди и таким образом снизить время ожидания пользователем результатов расчетов.

Существенной проблемой при организации глобальной очереди заданий является обеспечение бинарной переносимости исполняемых модулей пользовательских приложений [25]. Для решения этой проблемы в МСЦ РАН ведутся интенсивные исследования в области контейнерной виртуализации. Найденные методы и способы [26] представления пользовательских заданий в виде контейнеров могут быть применены при организации перераспределения заданий глобальной очереди в сети СКЦ. Развитие технологий виртуализации в настоящее время позволяет говорить о программно-определяемой инфраструктуре вычислительных центров, в которой виртуализованные ключевые подсистемы (вычислительная, сетевая и хранения данных) предоставляются пользователям в виде сервисов с заданным качеством. Показано [27], что реализация такой инфраструктуры позволяет обеспечить возможность каждому пользователю продуктивно решать свои задачи за приемлемое время с приемлемым уровнем затрат.

Немаловажным вопросом при построении системы управления является разработка алгоритмов глобального планирования заданий. Проведенные исследования [28] показали, что в случае применения децентрализованной схемы управления с равноправными диспетчерами хорошо работают аукционные методы планирования. Для случая абсолютных приоритетов заданий наибольшую эффективность демонстрирует алгоритм планирования, основанный на методе английского аукциона [28].

Децентрализованная схема управления позволяет поддерживать независимость отдельных СКЦ из состава сети, что соответствует существующей организационной структуре: центрами владеют и управляют разные научные организации в разных ведомствах (Минобрнауки России, МГУ им. М.В. Ломоносова, НИЦ «Курчатовский институт»). Негативным следствием децентрализации управления является необходимость регистрации каждого пользователя в каждом СКЦ, где ведется независимая собственная база (реестр) пользователей. В каждом СКЦ возникает непрофильная для научной организации проблема обработки и защиты персональных данных пользователей. В подавляющем большинстве случаев персональные данные пользователя требуются только для экстренной связи с ним, а также для формирования агрегированной обезличенной статистики (сколько молодых исследователей используют суперкомпьютерные ресурсы, каков процент пользователей с ученой степенью и т.п.). Подобного вида статистика без юридических и технических препятствий может быть запрошена суперкомпьютерным центром у организации — работодателя пользователя.

Отметим, что распределение квот вычислительных ресурсов в СКЦ удобно осуществлять по иерархической схеме: вначале квоты выделяются организациям и научным проектам, а далее — группам пользователей и пользователям в рамках определенного проекта. Руководители проектов и организаций заинтересованы в возможности самостоятельного перераспределения квот на ресурсы между своими проектами и пользователями.

Приемлемым решением видится создание распределенной системы аутентификации, авторизации и учета, в которой пользователь будет иметь единственную учетную запись, а институты могли бы гибко управлять выделяемыми или приобретаемыми суперкомпьютерными ресурсами. Подходящей основой для такой системы может стать удостоверяющая федерация [29] из суперкомпьютерных центров/институтов, доверяющих друг другу аутентификацию пользователя. Разрабатываемое в МСЦ РАН решение по построению удостоверяющей федерации сети СКЦ подробно рассмотрено в [30].

Единая система мониторинга предназначена для оперативного сбора информации о состоянии и загруженности сети СКЦ, ее хранения и отображения. Среди основных функций системы мониторинга следует отметить отслеживание состава сети СКЦ, характеристик входящих в состав суперкомпьютерных ресурсов; отображение текущей загруженности суперкомпьютерных ресурсов; мониторинг содержания выполняемой вычислительной работы (какие пользователи, где, и какие задания выполняют); отслеживание и отображение объемов израсходованных ресурсов, текущего состояния бюджета, выделенных квот ресурсов, настроек приоритетов, графиков изменения этих характеристик по всем научным проектам; мониторинг состояния вычислительных узлов суперкомпьютеров и каналов связи.

Проектом создаваемой сети СКЦ предусматривается выделение отдельного уровня отказоустойчивого хранения и глобального доступа к данным. На этом уровне пользователям предоставляется общее пространство хранения данных, доступное из любого СКЦ. Для организации общего пространства часть ресурсов систем хранения данных СКЦ использу-

ется для развертывания программно-определяемого хранилища данных — узла облачной системы хранения. На объединяемых узлах облачной системы хранения реализуется единое пространство имен файлов [31]. Доступ пользователей к данным организуется через шлюз облачной системы хранения [32], через который реализуются механизмы кэширования данных и политики автоматического перемещения «остывших» данных в облачную систему хранения. Надежность хранения данных в облачной системе хранения обеспечивается либо репликацией данных, либо использованием алгоритмов рассредоточения информации [33] между узлами облачной системы хранения.

4. Требования к телекоммуникационной инфраструктуре НИКС и сервисам на ее основе со стороны распределенной сети суперкомпьютерных центров коллективного пользования

Процесс выполнения высокопроизводительных вычислений можно разделить на три основных этапа:

– этап подготовки исходных данных;
– этап расчетов;
– этап анализа результатов и их визуализации.

На каждом из этапов задействуется свой вид вычислительной техники. Для подготовки данных требуется многопроцессорный сервер с большим объемом оперативной памяти (в настоящее время — несколько терабайт). Особая важность этапа подготовки (предобработки) данных заключается в том, что от точности исходных расчетных сеток зачастую зависит скорость и точность высокопроизводительных расчетов.

На этапе расчетов задействуются суперкомпьютерные ресурсы. Как правило, этот этап занимает наибольшее количество времени и вычислительных ресурсов.

На этапе анализа результатов и визуализации (постобработки) требуются рабочие станции, оснащенные развитой графической подсистемой.

В распределенной сети СКЦ серверы предобработки, суперкомпьютеры и станции постобработки могут принадлежать различным организациям и находиться на значительном территориальном удалении друг от друга. Поскольку связующим звеном между названными этапами высокопроизводительных вычислений являются пользовательские данные, для их автоматического и прозрачного для пользователя перемещения от серверов предобработки к суперкомпьютерам, от суперкомпьютеров к станциям постобработки, необходима единая для всей сети распределенная файловая система.

Отсутствие единого файлового пространства усложнит процесс управления вычислениями и распределения вычислительной нагрузки между суперкомпьютерами, а главное, усложнит работу пользователей.

Организация и функционирование такой файловой системы требует наличия надежной сетевой инфраструктуры, обеспечивающей гарантированные пропускную способность, низкую латентность и быстрое время отклика.

Рассмотрим минимальные требования к гарантированной пропускной способности телекоммуникационной сети для обеспечения работоспособности сети СКЦ. Размер исходных данных для расчетов и их результатов для одного задания в среднем составляет 100 Гбайт. Статистика МСЦ РАН показывает, что за год суперкомпьютеры центра выполняют около 100 тыс. заданий, т.е. средняя интенсивность входного потока — 1 задание в 5 минут.

Согласно результатам моделирования [28] перераспределению (перемещению между СКЦ) подвергается около 50% всех заданий, т.е. 1 задание в 10 минут.

Для перемещения в течение этого времени задания размером 100 Гбайт с учетом накладных расходов (заголовки кадров Ethernet и служебная информация) потребуется сеть передачи данных, обеспечивающая постоянную скорость передачи данных минимальной пропускной способностью 1,4 Гбит/с. Учитывая неравномерность передачи данных и особенности передачи данных на большие дистанции [34], только для передачи служебного трафика, вызванного перераспределеним задач, с уровнем задержек, не вызывающим перебои функционирования, потребуется пропускная способность на порядок больше — 10–15 Гбит/с. С учетом трафика удаленно работающих пользователей (загрузки и выгрузки данных в системы хранения сети СКЦ из научных организаций) требуемая пропускная способность возрастает до 15–20 Гбит/с.

В ближайшие три–пять лет следует ожидать рост производительности суперкомпьютерных систем в 2–3 раза, что приведет к росту детализации расчетных сеток и, как результат, к росту объема исходных данных и результатов в 2–3 раза, а следовательно, к росту требований к пропускной способности телекоммуникационной сети, необходимой для нормального функционирования сети СКЦ до 40 Гбит/с и выше.

Требования к пропускной способности телекоммуникационной сети можно снизить, если возможно предсказать гарантированное время доставки данных пользователя, и учесть его в процессе планирования распределения задач. Это возможно, если телекоммуникационная сеть будет иметь механизм выделения гарантированной полосы пропускания по требованию или по расписанию и предоставит планировщикам заданий сети СКЦ прикладной программный интерфейс к этому механизму.

Дополнительные требования по функциональности возникают при использовании контейнерного представляения заданий. Например, если исполняемое в интерактивном режиме задание получит дополнительные ресурсы (вычислительные узлы) на удаленной машине, связь с ними может быть организована через виртуальную частную сеть, которую необходимо создать средствами телекоммуникационной сети. Для этого телекоммуникационная сеть должна предоставить планировщику заданий сети СКЦ сответствующий прикладной программный интерфейс управления виртуальными частными сетями.

Упомянутые выше сервисы автоматизированного выделения сетевых ресурсов с использованием прикладных программных интерфейсов разрабатываются и уже предоставляются некоторыми зарубежными сетями науки и образования. Такие сервисы планируются и в национальной исследовательской компьютерной сети нового поколения.

Для функционирования сети СКЦ необходим непрерывный мониторинг как сетевой связности, так и характеристик обеспечиваемого качества связи. Национальные научно-образовательные сети обеспечивают такой мониторинг конечному пользователю на основе инструмента PerfSONAR [35]. Внедрение этого инструмента в НИКС является одной из приоритетных задач.

Для перераспределения задания пользователя с одного вычислительного ресурса сети СКЦ на другой необходимо поддерживать соответствие между учетными записями пользователя на этих ресурсах. Эта задача значительно упрощается при использовании единой учетной записи на всех ресурсах. Одним из способов реализации единой учетной записи является федеративная авторизация и аутентификация, при которой служба авторизации ресурса перенаправляет пользователя для аутентификации в организацию, в которой он

работает или учится. Сервисы федеративной аутентификации и авторизации, интегрированные в международные удостоверяющие федерации научно-образовательных сетей, уже поддерживаются в нашей стране сетями RASNet и RUNNet, а в рамках создаваемой на основе этих сетей НИКС планируется их дальнейшее развитие.

Заключение

Для полной реализации функциональности и обеспечения эффективности распределенной сети СКЦ необходима телекоммуникационная сеть со следующими требованиями:

— пропускная способность до 15–20 Гбит/с (с наращиванием в ближайшие 3–5 лет до 40 Гбит/с);
— автоматизированное выделение гарантированной полосы пропускания по требованию и по расписанию с использованием прикладного программного интерфейса;
— автоматизированное выделение виртуальных частных сетей по требованию и по расписанию с использованием прикладного программного интерфейса;
— средства сквозного мониторинга пути передачи данных с прикладным программным интерфейсом;
— поддержка федеративной аутентификации и авторизации.

В создаваемой национальной исследовательской компьютерной сети нового поколения планируется развитие функциональных возможностей, обеспечивающих выполение перечисленных выше требований, с учетом имеющегося опыта зарубежных национальных научно-образовательных сетей и интеграции российской национальной сети в мировое сообщество научно-образовательных сетей.

Публикация выполнена в МСЦ РАН в рамках государственного задания по проведению фундаментальных научных исследований.

Литература

1. Фортов В.Е., Савин Г.И., Левин В.К., Забродин А.В., Шабанов Б.М. Создание и применение системы высокопроизводительных вычислений на базе высокоскоростных сетевых технологий // Информационные технологии и вычислительные системы. 2002. № 1. С. 3.
2. Deutschen Forschungsnetz. URL: `https://www.dfn.de/` (дата обращения: 21.08.2019).
3. CANARIE. URL: `https://www.canarie.ca/` (дата обращения: 21.08.2019).
4. Internet2. URL: `https://www.internet2.edu/` (дата обращения: 21.08.2019).
5. SURFnet. URL: `https://www.surf.nl/en` (дата обращения: 21.08.2019).
6. AARNET. URL: `https://www.aarnet.edu.au/` (дата обращения: 21.08.2019).
7. China Educational and Research Network. URL: `http://www.edu.cn/english/` (дата обращения: 21.08.2019).
8. NORDUnet. Nordic gateway for Research and Education. URL: `https://www.nordu.net/` (дата обращения: 21.08.2019).
9. GÉANT. URL: `https://www.geant.org/` (дата обращения: 21.08.2019).
10. Asi@Connect. URL: `http://www.tein.asia` (дата обращения: 21.08.2019).
11. Asia Pacific Advanced Network. URL: `https://apan.net/` (дата обращения: 21.08.2019).

12. RedCLARA. Latin American Cooperation of Advanced Networks. URL: `https://www.redclara.net/` (дата обращения: 21.08.2019).

13. AfricaConnect2. URL: `https://www.africaconnect2.net/` (дата обращения: 21.08.2019).

14. Catlett C. The philosophy of TeraGrid: building an open, extensible, distributed TeraScale facility. Cluster Computing and the Grid 2nd IEEE/ACM International Symposium CCGRID2002, 2002. DOI: 10.1109/CCGRID.2002.1017101.

15. XSEDE — The Extreme Science and Engineering Discovery Environment. URL: `https://www.xsede.org/` (дата обращения: 21.08.2019).

16. Bassini S., Cavazonni C., Gheller C. European actions for High-Performance Computing: PRACE, DEISA and HPC-Europa. Il Nuovo Cimento C. 2009. Vol. 32. P. 93–97.

17. PRACE — Partnetship for Advanced Computing in Europe. URL: `http://www.prace-ri.eu/` (дата обращения: 21.08.2019).

18. Matsuoka S., Shimojo S., Aoyagi M., Sekiguchi S., Usami H., Miura K. Japanese Computational Grid Research Project: NAREGI. Proceedings of the IEEE. 2005. Vol. 93, no. 3. P. 522–533. DOI: 10.1109/JPROC.2004.842748.

19. PRACE: Europe's supercomputing infrastructure relies on GÉANT. URL: `https://impact.geant.org/portfolio/prace/` (дата обращения: 21.08.2019).

20. MD-VPN Product Description. URL: `https://wiki.geant.org/display/PLMTES/MD-VPN+Product+Description` (дата обращения: 21.08.2019).

21. XSEDE System Requirements Specification v3.1. URL: `http://hdl.handle.net/2142/45102` (дата обращения: 21.08.2019).

22. Шабанов Б.М., Овсянников А.П., Баранов А.В., Лещев С.А., Долгов Б.В., Дербышев Д.Ю. Проект распределенной сети суперкомпьютерных центров коллективного пользования // Программные системы: теория и приложения. 2017. № 4(35). С. 245–262. DOI: 10.25209/2079-3316-2017-8-4-245-262.

23. Шабанов Б.М., Телегин П.Н., Овсянников А.П., Баранов А.В., Тихомиров А.И., Ляховец Д.С. Система управления заданиями распределенной сети суперкомпьютерных центров коллективного пользования // Труды научно-исследовательского института системных исследований Российской академии наук. 2018. Т. 8, № 6. С. 65–73. DOI: 10.25682/NIISI.2018.6.0009.

24. Баранов А.В., Тихомиров А.И. Методы и средства организации глобальной очереди заданий в территориально распределенной вычислительной системе // Вестник Южно-Уральского государственного университета. Серия: Вычислительная математика и информатика. 2017. Т. 6, № 4. С. 28–42. DOI: 10.14529/cmse170403.

25. Шабанов Б.М., Телегин П.Н., Баранов А.В., Семенов Д.В., Чуваев А.В. Динамический конфигуратор виртуальной распределенной вычислительной среды // Программные продукты, системы и алгоритмы. 2017. № 4. DOI: 10.15827/2311-6749.25.272.

26. Baranov A.V., Savin G.I., Shabanov B.M. *et al.* Methods of Jobs Containerization for Supercomputer Workload Managers // Lobachevskii Journal of Mathematics. 2019. Vol. 40, no. 5. P. 525–534. DOI: 10.1134/S1995080219050020.

27. Шабанов Б.М., Самоваров О.И. Принципы построения межведомственного центра коллективного пользования общего назначения в модели программно-определяемого ЦОД

// Труды Института системного программирования РАН. 2018. Т. 30, № 6. С. 7–24. DOI: 10.15514/ISPRAS-2018-30(6)-1.

28. Baranov A., Telegin P., Tikhomirov A. Comparison of Auction Methods for Job Scheduling with Absolute Priorities // In: Malyshkin V. (eds) Parallel Computing Technologies (PaCT 2017). Lecture Notes in Computer Science. 2017. Vol. 10421. P. 387–395. DOI: 10.1007/978-3-319-62932-2_37.

29. Овсянников А.П., Савин Г.И., Шабанов Б.М. Удостоверяющие федерации научно-образовательных сетей // Программные продукты и системы. 2012. № 4. С. 3–7.

30. Баранов А.В., Овсянников А.П., Шабанов Б.М. Федеративная аутентификация в распределенной инфраструктуре суперкомпьютерных центров // Труды научно-исследовательского института системных исследований Российской академии наук. 2018. Т. 8, № 6. С. 79–83. DOI: 10.25682/NIISI.2018.6.0011.

31. Koulouzis S., Belloum A., Bubak M., Lamata P., Nolte D., Vasyunin D., de Laat C. Distributed Data Management Service for VPH Applications // IEEE Internet Computing. 2016. Vol. 20, no. 2. P. 34–41. DOI: 10.1109/MIC.2015.71.

32. Kapadia A., Varma S., Rajana K. Implementing Cloud Storage with OpenStack Swift. Packt Publishing, 2014. 105 p.

33. Джонс М. Анатомия облачной инфраструктуры хранения данных. Модели, функции и внутренние детали. 2012. URL: `https://www.ibm.com/developerworks/ru/library/cl-cloudstorage/cl-cloudstorage-pdf.pdf` (дата обращения: 28.08.2019).

34. Баранов А.В., Вершинин Д.В., Дербышев Д.Ю., Долгов Б.В., Лещев С.А., Овсянников А.П., Шабанов Б.М. Об эффективности использования канала связи между территориально удаленными суперкомпьютерными центрами // Труды научно-исследовательского института системных исследований Российской академии наук. 2017. Т. 7, № 4. С. 137–142.

35. Hanemann A. *et al.* PerfSONAR: A Service Oriented Architecture for Multi-domain Network Monitoring // Lecture Notes in Computer Science. 2005. Vol. 3826. P. 241–254. DOI: 10.1007/11596141_19.

Савин Геннадий Иванович, академик РАН, д.ф.-м.н., профессор, научный руководитель Межведомственного суперкомпьютерного центра РАН — филиала ФГУ ФНЦ НИИСИ РАН (Москва, Российская Федерация)

Шабанов Борис Михайлович, к.т.н., доцент, директор Межведомственного суперкомпьютерного центра РАН — филиала ФГУ ФНЦ НИИСИ РАН (Москва, Российская Федерация)

Баранов Антон Викторович, к.т.н., доцент, в.н.с. Межведомственного суперкомпьютерного центра РАН — филиала ФГУ ФНЦ НИИСИ РАН (Москва, Российская Федерация)

Овсянников Алексей Павлович, в.н.с. Межведомственного суперкомпьютерного центра РАН — филиала ФГУ ФНЦ НИИСИ РАН (Москва, Российская Федерация)

Гончар Андрей Андреевич, в.н.с. Межведомственного суперкомпьютерного центра РАН — филиала ФГУ ФНЦ НИИСИ РАН (Москва, Российская Федерация)

DOI: 10.14529/cmse200102

ON THE USE OF FEDERAL SCIENTIFIC TELECOMMUNICATION INFRASTRUCTURE FOR HIGH PERFORMANCE COMPUTING

© 2020 G.I. Savin, B.M. Shabanov, A.V. Baranov, A.P. Ovsyannikov, A.A. Gonchar

Joint SuperComputer Center of the Russian Academy of Sciences — Branch of Federal State Institution "Scientific Research Institute for System Analysis of the Russian Academy of Sciences" (Leninsky Prospect 32a, Moscow, 119334 Russia)

E-mail: savin@jscc.ru, shabanov@jscc.ru, antbar@mail.ru, ovsiannikov@jscc.ru, andrey.gonchar@jscc.ru

Received: 09.01.2020

The article is devoted to the prospects for the development of scientific telecommunications infrastructure based on the new generation national research computer network (NRCN), formed by the integration of departmental scientific and educational networks RUNNet and RASNet. The new network′ capabilities for combining supercomputer resources and providing barrier-free access to them are shown. Based on the generalized world experience, it has been shown that supercomputer infrastructures have special requirements for a telecommunication network for data transmission and the presence of a number of additional services. These requirements go far beyond the services of commercial telecom providers and, as a rule, can only be satisfied by the combined efforts of national scientific and educational networks.

The key elements of the federal telecommunications infrastructure necessary for combining high-performance computing resources are considered: high-performance communication channels with a specified quality of service, their automatic allocation on demand and on schedule, trusted network environment, federated authentication and authorization, reliability and security, end-to-end monitoring of the data transmission path between end users. Based on the analysis of the life cycle of the supercomputer job migrating at the distributed network, the requirements for the NRCN telecommunications infrastructure and services based on it are formulated.

Keywords: national science and education network, supercomputer center, shared research facilities, distributed computing, telecommunications infrastructure.

FOR CITATION

Savin G.I., Shabanov B.M., Baranov A.V., Ovsyannikov A.P., Gonchar A.A. On the Use of Federal Scientific Telecommunication Infrastructure for High Performance Computing. *Bulletin of the South Ural State University. Series: Computational Mathematics and Software Engineering.* 2020. Vol. 9, no. 1. P. 20–35. (in Russian) DOI: 10.14529/cmse200102.

References

1. Fortov V.E., Savin G.I., Levin V.K., Zabrodin A.V., Shabanov B.M. Creation and application of a high-performance computing system based on high-speed network technologies. Journal of Information Technologies and Computing. 2002. no. 1. P. 3. (in Russian)
2. Deutschen Forschungsnetz. Available at: `https://www.dfn.de/` (accessed: 21.08.2019).
3. CANARIE. Available at: `https://www.canarie.ca/` (accessed: 21.08.2019).
4. Internet2. Available at: `https://www.internet2.edu/` (accessed: 21.08.2019).

5. SURFnet. Available at: https://www.surf.nl/en (accessed: 21.08.2019).

6. AARNET. Available at: https://www.aarnet.edu.au/ (accessed: 21.08.2019).

7. China Educational and Research Network. Available at: http://www.edu.cn/english/ (accessed: 21.08.2019).

8. NORDUnet. Nordic gateway for Research and Education. Available at: https://www.nordu.net/ (accessed: 21.08.2019).

9. GÉANT. Available at: https://www.geant.org/ (accessed: 21.08.2019).

10. Asi@Connect. Available at: http://www.tein.asia (accessed: 21.08.2019).

11. Asia Pacific Advanced Network. Available at: https://apan.net/ (accessed: 21.08.2019).

12. RedCLARA. Latin American Cooperation of Advanced Networks. Available at: https://www.redclara.net/ (accessed: 21.08.2019).

13. AfricaConnect2. Available at: https://www.africaconnect2.net/ (accessed: 21.08.2019).

14. Catlett C. The philosophy of TeraGrid: building an open, extensible, distributed TeraScale facility. Cluster Computing and the Grid 2nd IEEE/ACM International Symposium (CCGRID 2002). 2002. DOI: 10.1109/CCGRID.2002.1017101.

15. XSEDE — The Extreme Science and Engineering Discovery Environment. Available at: https://www.xsede.org/ (accessed: 21.08.2019).

16. Bassini S., Cavazonni C., Gheller C. European actions for High-Performance Computing: PRACE, DEISA and HPC-Europa. Il Nuovo Cimento C. 2009. Vol. 32. P. 93–97.

17. PRACE — Partnetship for Advanced Computing in Europe. Available at: http://www.prace-ri.eu/ (accessed: 21.08.2019).

18. Matsuoka S., Shimojo S., Aoyagi M., Sekiguchi S., Usami H., Miura K. Japanese Computational Grid Research Project: NAREGI. Proceedings of the IEEE. 2005. Vol. 93, no. 3. P. 522–533. DOI: 10.1109/JPROC.2004.842748.

19. PRACE: Europe's supercomputing infrastructure relies on GÉANT. Available at: https://impact.geant.org/portfolio/prace/ (accessed: 21.08.2019).

20. MD-VPN Product Description. Available at: https://wiki.geant.org/display/PLMTES/MD-VPN+Product+Description (accessed: 21.08.2019).

21. XSEDE System Requirements Specification v3.1. Available at: http://hdl.handle.net/2142/45102 (accessed: 21.08.2019).

22. Shabanov B., Ovsiannikov A., Baranov A., Leshchev S., Dolgov B., Derbyshev D. The distributed network of the supercomputer centers for collaborative research. Program systems: Theory and applications. 2017. no. 8:4(35). P. 245–262. (in Russian) DOI: 10.25209/2079-3316-2017-8-4-245-262.

23. Shabanov B.M., Telegin P.N., Ovsyannikov A.P., Baranov A.V., Tikhomirov A.I., Lyakhovets D.S. The Jobs Management System for the Distributed Network of the Supercomputer Centers. The Proceeding of the Scientific Research Institute for System Analysis of the Russian Academy of Sciences. 2018. Vol. 8, no. 6. P. 65–73. (in Russian) DOI: 10.25682/NIISI.2018.6.0009.

24. Baranov A.V., Tikhomirov A.I. Methods and Tools for Organizing the Global Job Queue in the Geographically Distributed Computing System. Bulletin of the South Ural State University. Series: Computational Mathematics and Software Engineering. 2017. Vol. 6, no. 4. P. 28–42. (in Russian) DOI: 10.14529/cmse170403.

25. Shabanov B.M., Telegin P.N., Baranov A.V., Semenov D.V., Chuvaev A.V. Dynamic Configurator for Virtual Distributed Computing Environment. Software Journal: Theory and Applications. 2017. no. 4. (in Russian) DOI: 10.15827/2311-6749.25.272.

26. Baranov A.V., Savin G.I., Shabanov B.M. *et al.* Methods of Jobs Containerization for Supercomputer Workload Managers. Lobachevskii Journal of Mathematics. 2019. Vol. 40, no. 5. P. 525–534. DOI: 10.1134/S1995080219050020.

27. Shabanov B.M., Samovarov O.I. Building the Software Defined Data Center. Proceedings of the Institute for System Programming. 2018. Vol. 30, no. 6. P. 7–24. (in Russian) DOI: 10.15514/ISPRAS-2018-30(6)-1.

28. Baranov A., Telegin P., Tikhomirov A. Comparison of Auction Methods for Job Scheduling with Absolute Priorities. In: Malyshkin V. (eds) Parallel Computing Technologies (PaCT 2017). Lecture Notes in Computer Science. 2017. Vol. 10421. P. 387–395. DOI: 10.1007/978-3-319-62932-2_37.

29. Ovsyannikov A.P., Savin G.I., Shabanov B.M. Identity federation of the research and educational networks. Software & Systems. 2012. no. 4. P. 3–7. (in Russian)

30. Baranov A.V., Shabanov B.M., Ovsyannikov A.P. Federative Identity for the Distributed Infrastructure of the Supercomputer Centers. The Proceeding of the Scientific Research Institute for System Analysis of the Russian Academy of Sciences. 2018. Vol. 8, no. 6. P. 79–83. (in Russian) DOI: 10.25682/NIISI.2018.6.0011.

31. Koulouzis S., Belloum A., Bubak M., Lamata P., Nolte D., Vasyunin D., de Laat C. Distributed Data Management Service for VPH Applications. IEEE Internet Computing. 2016. Vol. 20, no. 2. P. 34–41. DOI: 10.1109/MIC.2015.71.

32. Kapadia A., Varma S., Rajana K. Implementing Cloud Storage with OpenStack Swift. Packt Publishing, 2014. 105 p.

33. Jones M. Anatomy of a cloud storage infrastructure. Models, features, and internals. 2010. Available at: `https://developer.ibm.com/articles/cl-cloudstorage/` (accessed: 21.08.2019).

34. Baranov A.V., Derbyshev D.Yu., Dolgov B.V., Leshchev S.A., Ovsyannikov A.P., Shabanov B.M., Vershinin D.V. Effective usage of the link between geographically distributed supercomputer centers. The Proceeding of the Scientific Research Institute for System Analysis of the Russian Academy of Sciences. 2017. Vol. 7, no. 4. P. 137–142. (in Russian)

35. Hanemann A. *et al.* PerfSONAR: A Service Oriented Architecture for Multi-domain Network Monitoring. Lecture Notes in Computer Science. 2005. Vol. 3826. P. 241–254. DOI: 10.1007/11596141_19.