

=====ИНФОРМАЦИОННЫЕ И ВЫЧИСЛИТЕЛЬНЫЕ =====
=====ТЕХНОЛОГИИ В БИОЛОГИИ И МЕДИЦИНЕ=====

УДК: 004.77:004.62:004.9

Перспективы использования центров обработки данных при решении задач математической биологии и биоинформатики

Корнилов В.В.^{*1,2}, Исаев Е.А.^{1,3}, Исаев К.А.²

¹*Национальный исследовательский университет «Высшая школа экономики», Москва, 101000, Россия*

²*Институт математических проблем биологии, Российская академия наук, Пущино, Московская область, 142290, Россия*

³*Пушчинская Радиоастрономическая обсерватория Астрокосмического центра ФИАН, Пущино, Московская область, 142290, Россия*

Аннотация. В статье анализируется состояние современной «науки о жизни» - вычислительной биологии, рассматривается проблема «больших данных» и способы решения этой крайне актуальной для современной науки задачи, связанной с необходимостью хранения и оперативной аналитической обработки сверхбольших объёмов информации, получаемой в ходе научных экспериментов, анализируется потенциал центров обработки данных и, особенно, модульных центров обработки данных, как основного средства обеспечения научных центров необходимыми для успешного решения их исследовательских задач вычислительными ресурсами.

Ключевые слова: математическая биология, биоинформатика, большие данные, облачные вычисления, центры обработки данных, дата-центры, модульные центры обработки данных.

ВВЕДЕНИЕ

Проведение исследований на современном уровне в таких научных областях, как математическая биология, биоинформатика и смежных с ними тесно связано с новейшими достижениями в области информационных технологий и в основном базируется на обработке больших объёмов цифровой информации. Уменьшение стоимости лабораторных установок по секвенированию ДНК, повсеместное введение в практику методов математического моделирования, переход к персонализированной медицине и другие инновационные проекты приводят к необходимости хранения, обработки и анализа сверхбольших объёмов информации, в идеале – в реальном масштабе времени. Понятно, что для решения такой задачи необходимы не просто высокопроизводительные компьютерные ресурсы, а суперкомпьютеры, специализированные для обработки больших данных. Естественно, далеко не каждая исследовательская лаборатория может позволить себе иметь в своём распоряжении подобные установки, стоимость которых остаётся достаточно высокой. Выходом из сложившейся ситуации является использование коллективных центров обработки данных (ЦОД). Обычно ЦОД представляют собой отдельные сооружения со специальным оборудованием, включающим вычислительные сервера, телекоммуникационные устройства, системы охлаждения и средства обеспечения бесперебойного электропитания. Функциональное назначение ЦОД понятно уже из его

*basil@psn.ru

названия, они обеспечивают гарантированно безопасное хранение пользовательских данных, бесперебойный удалённый доступ к этим данным и их высокоскоростную обработку, предоставляя по требованию пользователей необходимые вычислительные ресурсы. Стоимость услуг ЦОД стабильно снижается, особенно с развитием концепции модульных ЦОД, имеющих существенно более низкую стоимость и быстрое время развёртывания по сравнению с классическими ЦОД. Для научных центров использование именно модульных центров обработки данных является наиболее рациональным решением в силу их дешевизны, скорости развёртывания и настройки и существенно более быстрой системой модернизации основных элементов серверного и сетевого оборудования, что также значительно повышает их конкурентоспособность и снижает стоимость услуг.

ВЫЧИСЛИТЕЛЬНАЯ БИОЛОГИЯ

Практически ни одна из областей современной науки сегодня не обходится без использования методов математического моделирования с активным применением средств вычислительной техники. Понятие «вычислительная наука» прочно закрепилось в практике современных научных исследований и сегодня применимо буквально к любой научной дисциплине: вычислительная математика, вычислительная химия, вычислительная физика, вычислительная механика, вычислительная биология и т.д. И это отнюдь не дань моде на использование компьютеров – технический прогресс в области вычислительных и коммуникативных технологий сделал возможным коренным образом изменить парадигму проведения научных исследований. Наряду с теорией и экспериментом, вычислительная наука сегодня образует третью основную составляющую научного метода познания, позволяя с помощью компьютерного эксперимента исследовать модели сложных явлений и получать качественно новые знания об окружающем мире. Интересен отчёт, выполненный Консультативным комитетом по информационным технологиям при Президенте США, полностью посвящённый концепции вычислительной науки и её роли на современном этапе научного познания [1]. В частности там утверждается, что «наиболее важные и экономически перспективные научные открытия XXI века будут совершены благодаря информационным технологиям и приложениям вычислительной науки». И далее: «несмотря на фундаментальный вклад вычислительной науки в новые открытия, безопасность и конкурентоспособность, неадекватные и устаревшие структуры Федерального Правительства и Академии (США) не достаточно поддерживают это критически важное междисциплинарное направление. Это несоответствие создает угрозу лидерству науки США, экономической конкурентоспособности и национальной безопасности». Пожалуй, сложно более ёмко и выразительно охарактеризовать роль вычислительных методов и алгоритмов компьютерного моделирования в современном научном исследовании.

Биология сегодняшнего дня также уже превратилась в вычислительную биологию, сложное междисциплинарное направление, использующее достижения информатики и вычислительной техники, прикладной математики и статистики для решения проблем, поставленных «наукой о жизни» [2, 3]. Всё разнообразие задач вычислительной биологии можно разделить на такие направления исследований, как математическое моделирование в биологии, широко использующее структурное моделирование и имитационные методы в исследовании, биоинформатику и вычислительную геномику, использующие машинные алгоритмы и статистические методы в анализе наборов биологических данных, состоящих, как правило, из большого числа ДНК, РНК и белковых последовательностей, в том числе, с применением методов геномного секвенирования. Реализация математических моделей осуществляется с помощью методов вычислительной математики, которая непрерывно совершенствуется вместе с прогрессом в области компьютерной техники. Теоретическое исследование сложных

процессов, допускающих математическое описание, осуществляется посредством вычислительного эксперимента. Используются методы численного решения систем линейных алгебраических уравнений, методы решения задачи Коши и краевых задач для линейных систем дифференциальных уравнений в частных производных, методы численного решения интегро-дифференциальных и интегральных уравнений. Биоинформатика активно использует математический аппарат различных дисциплин – вычислительной математики, теории вероятностей и математической статистики, информатики, теории алгоритмов, методов оптимизации и распознавания образов, теории искусственного интеллекта, теории графов, теории групп и др. Применяя этот аппарат, биоинформатика создает алгоритмы и программы, которые занимаются поиском новых знаний в большом потоке гетерогенной информации.

Использование современных информационных технологий является неременным условием для успешного решения задач, возникающих при изучении сложных биологических объектов. Действительно, массовое использование исследователями высокопроизводительных персональных компьютеров, растущая доступность высокопроизводительных вычислительных систем (специализированных суперкомпьютеров и мощных вычислительных кластеров), возможность использования систем распределённых вычислений, таких как сети GRID [5], дают современным биологам совершенно новые возможности по постановке модельных экспериментов, значительно сокращают время получения результатов этих экспериментов, позволяют решать задачи недостижимых ранее масштабов.

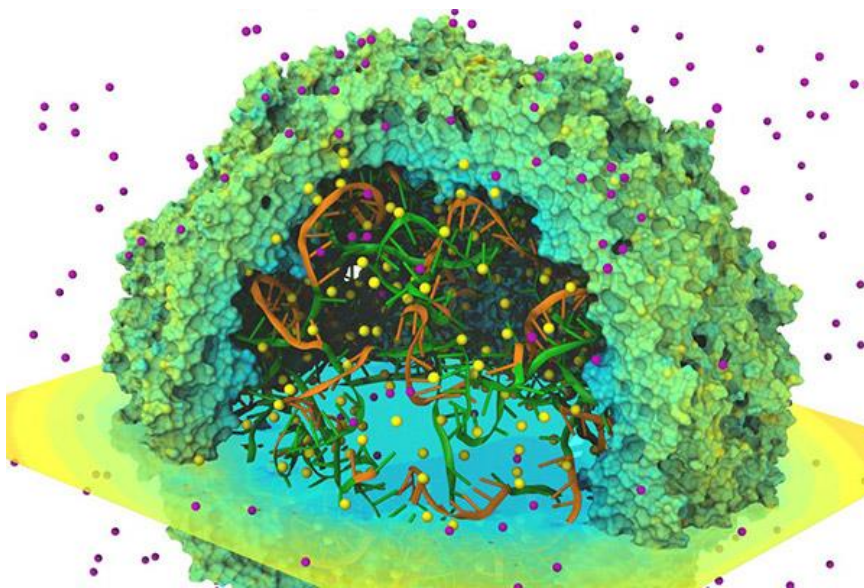


Рис. 1. Вирусная частица с нуклеиновой кислотой, белковой оболочкой и окружающими ионами в растворе (изображение: Theoretical and Computational Biophysics Group/University of Illinois at Urbana-Champaign).

К примеру, первоначальный проект по расшифровке генома человека, начатый в 1990 году, потребовал более 10 лет совместных усилий международных групп исследователей и более 5 млрд. долларов совокупных затрат. В настоящее время установки по секвенированию становятся рутинным инструментом многих научных лабораторий, тем самым приближая решение одной из основных задач персонализированной медицины – возможность узнать и проанализировать геномную информацию каждого конкретного пациента. В работах по молекулярному моделированию биологических объектов ещё вчера казались грандиозными задачи, в которых удавалось пронаблюдать за поведением молекулярных систем из нескольких тысяч атомов на протяжении нескольких наносекунд. А уже сегодня моделируется динамика сложнейших биологических объектов, состоящих из десятков и сотен тысяч

атомов на временах, перешагнувших миллисекундный рубеж (рис. 1). Такие количественные перемены приводят к значительным качественным изменениям научной деятельности, новым грандиозным открытиям.

Однако здесь мы сталкиваемся с проблемой больших данных (рис. 2). Широкое использование самых последних достижений информатики и вычислительной техники приводит к лавинообразному увеличению количества экспериментальных данных. Громадные физические объёмы информации, высокая скорость прироста новых данных и соответственно требующаяся высокая скорость их обработки, разнообразие форматов и структур данных при необходимости их одновременной обработки, в совокупности образуют новую парадигму обработки информации [6].

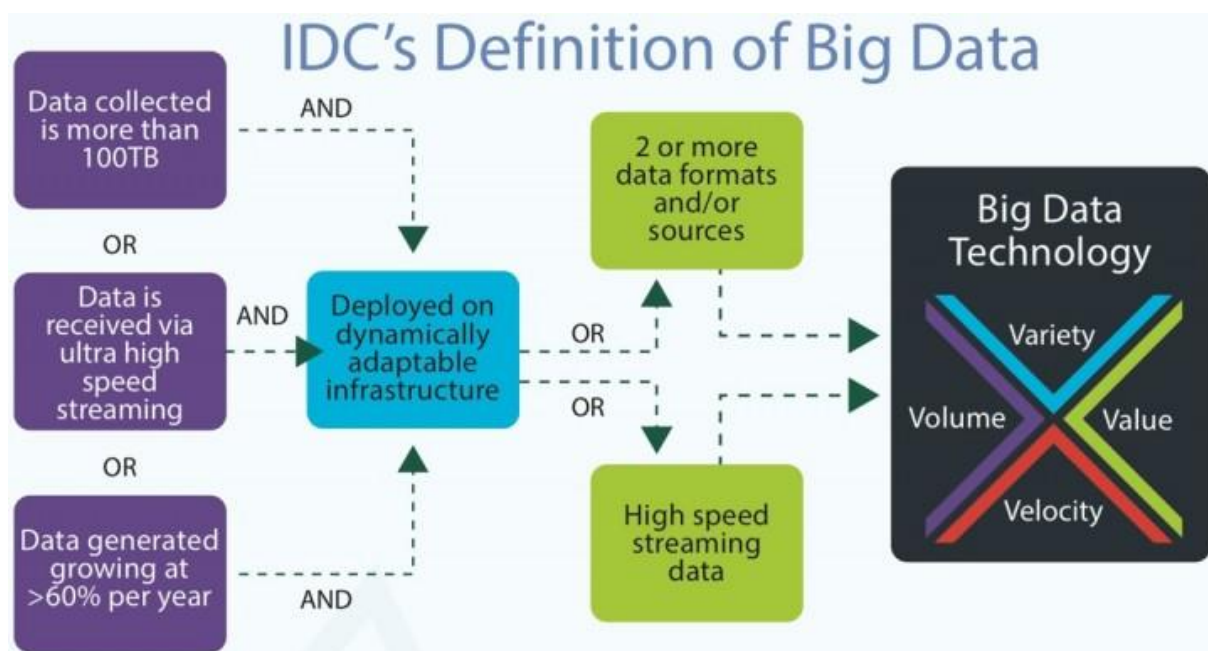


Рис. 2. Определение больших данных исследовательской компании IDC (изображение IDC).

Первоначальная запись и хранение сверхбольших массивов данных, их обработка и анализ, передача данных по информационным сетям – каждая из этих задач сложна, а часто и вовсе не разрешима с использованием традиционных средств и методов работы с данными. Одним из решений проблемы являются специализированные суперкомпьютеры для работы с большими данными, предоставляющие как возможности хранения больших объемов данных, в том числе распределенные, так и средства эффективного анализа данных в реальном масштабе времени. К сожалению, число предлагаемых к настоящему времени решений не слишком велико, а их стоимость далеко превосходит достаточно скромные финансовые возможности бюджетов большинства научно-исследовательских лабораторий. Некоторую помощь в борьбе с большими данными оказывают системы распределённых вычислений, такие как научные GRID-сети и вычислительные кластеры. Однако наилучшим решением зачастую оказывается использование концепции облачных вычислений, которая предполагает обеспечение удаленного динамического доступа пользователей к услугам, вычислительным ресурсам и приложениям (включая операционные системы и инфраструктуру) через существующие каналы передачи данных. Именно облачные технологии способны представить пользователям необходимые процессорные мощности для обработки сколь угодно больших объемов данных. Показателем недавно установленный рекорд производительности, установленный объединённым кластером, принадлежащим одному из лидеров в области предоставлении услуг облачных вычислений компании Amazon (рис. 3). Пиковая мощность виртуального облачного компьютера под управлением программной системы CycleServer компании Cycle

Computing в течение восемнадцати часов превышала 1,21 Пфлопс, а на каждую виртуальную машину приходилось в среднем по 9,3 процессорного ядра [7]. Такой результат позволяет сравнить его по производительности с суперкомпьютерами, занимающими лидирующие позиции в рейтинге Top-500 [8].

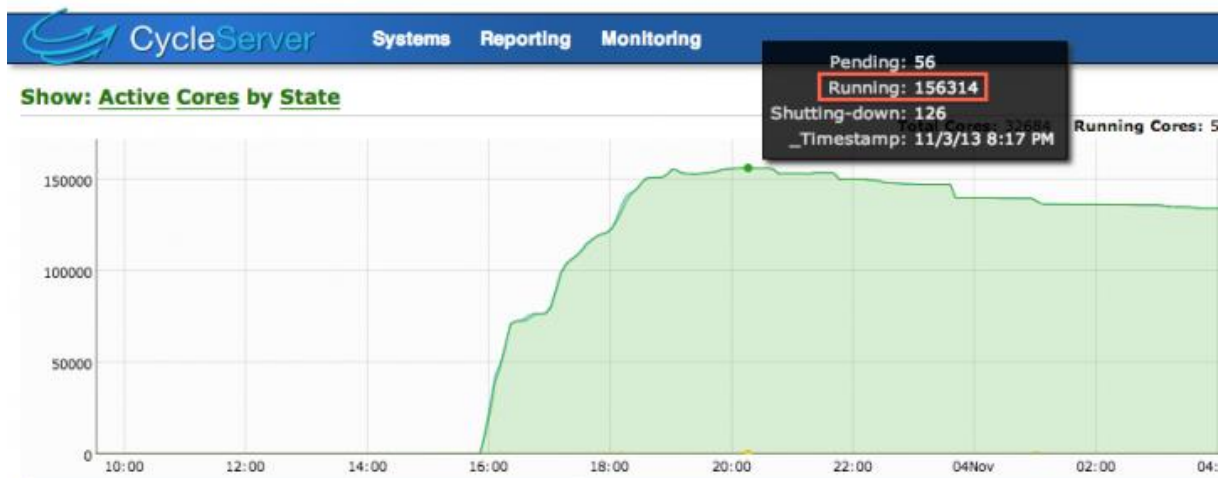


Рис. 3. Рекорд Cycle Computing, зарегистрированный на управляющем сервере [4].

Однако, назначение облачных вычислений вовсе не в достижении пиковой производительности при решении единственной вычислительной задачи. Их основные преимущества – удобство доступа, масштабируемость предоставляемых услуг, их относительно невысокая, по сравнению с приобретением суперкомпьютера или даже ценою его аренды, стоимость. Для подавляющего большинства научных учреждений сервисы облачных вычислений сегодня выглядят намного привлекательнее, чем использование суперкомпьютера любой классической архитектуры. Действительно, получить возможность загрузить одну из таких машин своими вычислительными задачами крайне сложно. Надо обосновать важность решаемой задачи и дождаться своей очереди. Затем требуется переписать свой код вместе с программистами обслуживающего суперкомпьютер института и полностью оплатить все время его использования, независимо от реального объема вычислений. Облачные платформы доступны сразу, легко масштабируются и обходятся значительно дешевле.

Не останавливаясь в данной работе на программно-алгоритмических методах реализации облачных вычислений, обратимся к технической стороне построения облаков. С точки зрения пользователя он сразу получает в свое распоряжение виртуальный сервер необходимой ему вычислительной мощности. При этом провайдеры облачных услуг принимают на себя обязанности по решению комплекса сложных задач по технической поддержке вычислительных серверов, обеспечению их бесперебойной работы, организации систем связи, динамического перераспределения вычислительных мощностей и многие другие. Наиболее грамотным решением является сосредоточение вычислительных серверов, систем хранения данных и обеспечивающих систем в комплексных решениях, известных как центры обработки данных.

ЦЕНТРЫ ОБРАБОТКИ ДАННЫХ

Итак, для хранения и обработки большого количества информации необходимо использовать специализированные технические решения, мощные серверы, дисковые хранилища. Создавать и обслуживать такие технические системы достаточно сложно и дорого: содержание серверов требует специальных технических условий, отдельных помещений и квалифицированного персонала. Одним из основных назначений центров

обработки данных как раз и является создание подходящих условий для размещения таких технических решений.

Согласно общепринятому определению, центры обработки данных (ЦОД, они же: data-center, дата-центры) – это вычислительная инфраструктура (набор взаимосвязанных программных и аппаратных компонентов, организационных процедур, мест размещения и персонала), предназначенная для безопасной централизованной обработки, хранения и предоставления данных, сервисов, приложений и обеспечивающая высокую степень виртуализации своих ресурсов (рис. 4). К основным задачам ЦОД в первую очередь относятся эффективное консолидированное хранение и обработка данных, предоставление пользователям прикладных сервисов, а также поддержка функционирования внешних приложений.

Перемещение разрозненных вычислительных ресурсов в единый ЦОД позволяют обеспечить комфортные условия для оборудования и технического персонала, а также увеличить степень физической защиты серверов. При использовании ЦОД можно снизить расходы на техническую поддержку и управление вычислительными ресурсами. Централизация также облегчает использование стандартизованных конфигураций и процессов управления, создание рентабельных систем резервного копирования и энергообеспечения. Упрощается и решение вопросов организации высококачественного контроля за состоянием окружающей среды и обеспечения физической защиты. Может быть улучшена и сетевая безопасность, поскольку серверы оказываются под защитой единого, централизованно управляемого межсетевого экрана.

Таким образом, дата-центры предназначены для размещения специализированных компьютерных устройств (вычислительные мощности, а также оборудование для хранения данных), используемых для хранения и высокопроизводительной обработки информации, а также на предоставлении клиентам удалённого доступа к информации по защищённым телекоммуникационным каналам передачи данных. При этом потребителям услуг дата-центра гарантируется бесперебойный доступ к своим данным.

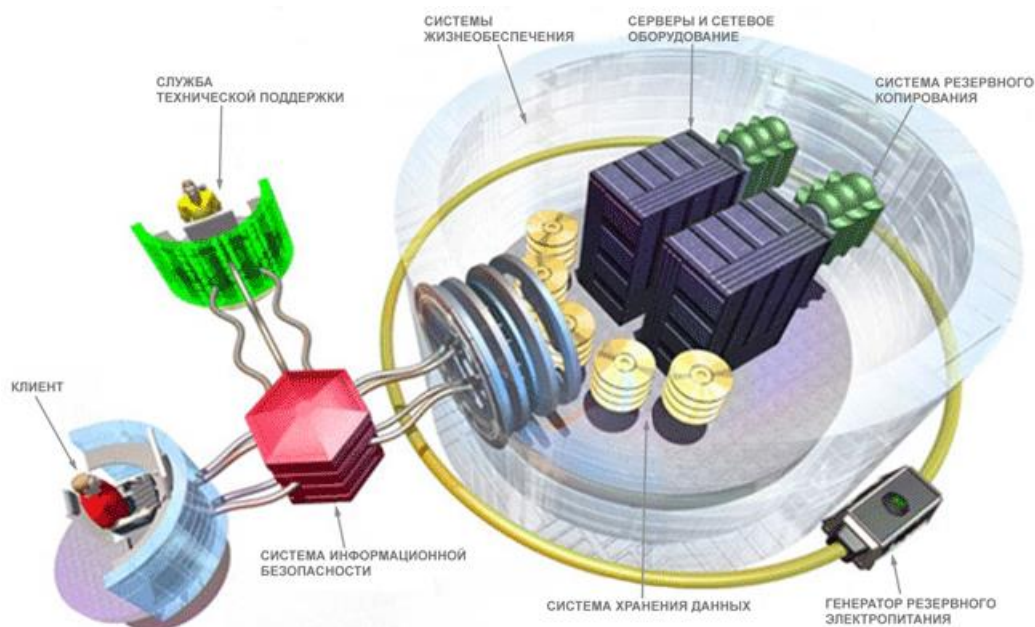


Рис. 4. Центр обработки данных: принципиальная схема (изображение RadiusGroup).

Преимущество дата-центра в том, что человек или компания, может из любой точки планеты, используя любое устройство доступа в сеть, работать с необходимой вычислительной мощностью, требуемым объемам памяти, всем необходимым

программным обеспечением, которое будет работать и храниться на серверах в дата-центре. Традиционные услуги в дата-центрах: аренда стойки, размещение серверов, подключение к Интернету, аренда каналов связи, установка, настройка ПО, администрирование. В настоящее время к ним добавились предоставление в аренду вычислительных мощностей, виртуальных серверов, дискового пространства для резервного копирования данных, аренды приложений.

К основным принципам функционирования современного ЦОД принято относить такие параметры, как виртуализация, под которой принято понимать агрегацию, распределение или добавление множества ресурсов в большие ресурсы или объединение ресурсов; кластеризация, обеспечивающая объединение нескольких серверов в единую систему, путем установки программных связей между ними, с целью координации их работы и выполнения перераспределения нагрузки в случае сбоя в одном из серверов; масштабирование, означающее возможность в процессе развития ЦОД добавлять к имеющимся новые вычислительные модули без замены существующего оборудования; и резервирование, залог того, что при выходе из строя одной подсистемы центра, функцию ее возьмет на себя другая, резервная.

Обязательные компоненты, входящие в состав ЦОД, можно разделить на три основные группы: технические компоненты, создающие условия для эффективной работы центра и включающие собственно серверы информационных ресурсов, приложений, представления информации, а также служебные серверы, систему хранения данных и резервного копирования, сетевую инфраструктуру, инженерную систему эксплуатации ЦОД и систему безопасности; программное обеспечение, включающее сервисы инфраструктуры ЦОД и ПО для корректной работы пользовательских приложений; а также среду организации процессов оказания услуг, в первую очередь их качество и доступность. Основным фактором, характеризующим качество услуг, предоставляемых дата-центром, является допустимое время незапланированного простоя оборудования, в идеале стремящееся к нулю. В качестве примера можно привести данные исследования Ponemon Institute [9]. Согласно данным этого исследования средняя стоимость одной минуты простоя дата-центров увеличилась с 2010 года на 41% и в 2013 году составила в США немногим более \$7,900 долларов. Среднее время полной неработоспособности ЦОД составило 119 минут, равняясь приблизительно \$901,500 общей стоимости.

Для выполнения требований, позволяющие гарантировать устойчивость ЦОД, а также минимизировать связанные с этим риски, учитывают специальные стандарты, регламентирующие параметры дата-центров. В частности стандарт «Tier Standard: Operational Sustainability», разработанный организацией Uptime Institute, устанавливает 4 уровня надёжности ЦОД, в зависимости от качества сервиса, предоставляемого ЦОД и уровня затрат на создание соответствующей инфраструктуры [10]. Особенность данного стандарта в том, что он учитывает человеческий фактор в устойчивой работе дата-центров. Также в соответствии со стандартом для устойчивости дата-центров имеют значение такие факторы, как расположение дата центра, его удаленность от пересечения транспортных путей – аэропортов, ж/д вокзалов, автодорог, особенности здания, в котором расположен ЦОД и уровень его энергоэффективности.

При анализе параметров, характеризующих дата-центры, бросается в глаза отсутствие чётких критериев размеров ЦОД. В разных источниках под ЦОД может подразумеваться и небольшое помещение, оборудованное под размещение всего нескольких серверов, и «мега центры обработки данных», занимающие площади в десятки тысяч квадратных метров и содержащие тысячи серверных стоек [11]. При этом, согласно аналитическому прогнозу от IDC (International Data Corporation) [12], число центров обработки данных будет продолжать расти, пока не достигнет максимума в 8,6 млн. к 2017 году, после чего начнётся их спад. При этом, так как ИТ-составляющая в успешной компании должна являться стратегически важным фактором снижения затрат, а не дополнительной статьёй расходов, прогнозируется, что в течении

ближайших пяти лет большая часть организаций откажется от поддержки собственной серверной инфраструктуры и будет стремиться передать управление их данными поставщикам соответствующих услуг. При этом от провайдеров потребуются переход на качественно новый уровень эксплуатации ЦОД и предоставления сервиса. Последнее утверждение является особенно актуальным, так как вопросы сохранности данных, информационной безопасности, обеспечения бесперебойного доступа к данным, становятся первоочередными при передаче функций по хранению и обработке данных «в чужие руки». Достаточно единственно инцидента с потерей критически важных для компании-клиента данных (причём для провайдера услуг ЦОД это событие может показаться весьма не значительным на фоне общих масштабов «мега ЦОД»), чтобы расходы на создание и поддержку собственного корпоративного дата-центра уже не будут выглядеть избыточно большими по сравнению с арендной платой за использование удалёнными серверами. Данный тезис особенно важен с учётом реалий российской действительности, где уровень предложения соответствующих услуг сравнительно небольшой, а качество сервиса, к сожалению, зачастую оставляет желать лучшего. В этой ситуации в нашей стране многие компании по-прежнему склоняются к необходимости создания собственных ЦОД.

Согласно аналитическому отчету iKS-Consulting [13], услуги коммерческих ЦОД в РФ предоставляют более 180 крупных и средних коммерческих дата-центров. Российский рынок дата-центров формируется крупными площадками, большая часть из которых расположена в Москве. 10 крупнейших операторов ЦОД составляют порядка 50% рынка по количеству введенных в эксплуатацию стоек (рис. 5). Все топ-10 операторов дата-центров России имеют московские площадки. В Москве и Московской области услуги ЦОД предоставляет порядка 110 компаний, общая площадь машинных залов которых составляет почти 60 тыс. кв.м (68,6% российского рынка по площадям машинных залов). В 2014 г. общее количество коммерческих стоек в дата-центрах Москвы выросло на 31% и составило 19,2 тыс.

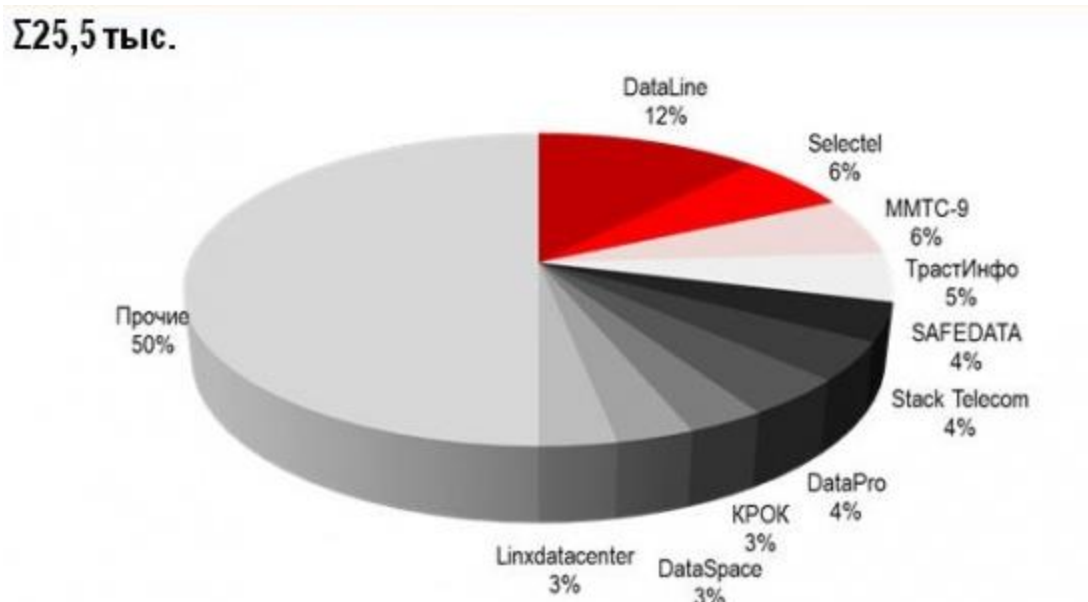


Рис. 5. Топ-10 коммерческих дата-центров России, предварительная статистика за 2014 год (источник iKS-Consulting).

Спрос на услуги ЦОД стабильно увеличивается, чему в немалой степени способствуют как общие тенденции – проблематика больших данных, растущая популярность облачных вычислений, так и новые законодательные инициативы в виде ожидаемого вступления в действие ФЗ №152-ФЗ «О персональных данных», требующего хранения персональных данных на территории РФ. Ожидается, что к 2018

году объем рынка вырастет почти в 2 раза и превысит 26,3 млрд, а количество установленных стоек в коммерческих дата-центрах увеличится до 48,3 тыс. [13].

Процесс создания классического ЦОД – сложный, многоступенчатый, весьма длительный по времени и очень дорогой по необходимым финансовым затратам. Обычно он начинается с этапа выбора площадки для строительства, отвечающей всем необходимым требованиям, включает этап собственно строительства здания или комплекса зданий ЦОД, этап реализации, заключающийся в комплектации и поставке необходимого оборудования, монтажу, внедрению основных компонентов, тестированию их, созданию организационной структуры, а также этапы опытной эксплуатации и модернизации. При этом известно, что новейшее компьютерное оборудование морально устаревает в срок до 5 лет. За это время ЦОД за счёт поставляемых услуг должен полностью окупить затраты на его создание, обслуживание, ремонт и модернизацию. В качестве примера приведём информацию о недавно начатом строительстве центра обработки данных в Подмоскowie по заказу компании «Авантаж». Сообщается, что общая ориентировочная стоимость создания этого дата-центра, который обещает стать крупнейшим российским ЦОД - он рассчитан на 2200 стоек с серверами, составит около 4 млрд. руб.

Итак, мы имеем базовое противоречие между насущной необходимостью использования высокопроизводительного компьютерного оборудования для обработки и хранения данных с одной стороны, значительной величиной арендной платы при использовании коммерческих ЦОД одновременно с достаточно проблематичной ситуацией с надёжностью и качеством этих услуг, с другой стороны. и фантастически высокой стоимостью начальных капиталовложений при создании собственного ЦОД. Ещё одна проблема, связанная с ЦОД в их классическом исполнении – наблюдающееся истощение ресурсов. Площадки московских ЦОД уже близки к заполнению, их операторы ищут возможности для расширения, но это не всегда просто из-за дефицита энергоёмкостей, свободных помещений и т. п.

Ещё более это актуально для научного сообщества. Не является тайной, что получить разовое финансирование на создание собственной вычислительной инфраструктуры для научно-исследовательского института может быть намного проще, чем изыскать возможности регулярной оплаты услуг аутсорсинга коммерческих дата-центров. Однако, разумеется, размер этого финансирования значительно (в десятки раз) меньше сумм, необходимых для создания традиционного ЦОД, а время его строительства превышает любые разумные сроки выполнения научно-исследовательского проекта.

Решением этих противоречий являются новые тенденции в создании дата-центров, а именно переход к технологии модульных центров обработки данных, то есть дата-центров, построенных из специальных, заранее изготовленных компонентов различной функциональной направленности.

МОДУЛЬНЫЕ ЦЕНТРЫ ОБРАБОТКИ ДАННЫХ

Модульный ЦОД (МЦОД) – это центр обработки данных, состоящий из типовых элементов системы – модулей. Каждый модуль выполняет различные функции – это могут быть модули, содержащие серверные блоки, модули со средствами энергоснабжения, системы охлаждения и т.д. МЦОД собирается из отдельных готовых модульных блоков, уже содержащих всё необходимое оборудование (рис. 6).

Рассмотрим ряд преимуществ МЦОД от классических дата-центров. Прежде всего, это значительные снижения денежных и временных затрат на строительство ЦОД. МЦОД не требуют капитального строительства или дорогостоящего переоборудования существующего помещения, практически отсутствуют расходы на проектирование и согласования строительства, что, в совокупности, как минимум, в половину удешевляет стоимость ЦОД и в несколько раз сокращает сроки ввода его в эксплуатацию. То же

самое относится и к оборудованию ЦОД вычислительными и инженерными системами. Создание модуля в заводских условиях намного проще и дешевле по сравнению с традиционными ЦОД, собираемыми по частям из различных устройств. Значительно снижается объём работ по подготовке инженерной инфраструктуры, а также уменьшаются затраты по проектированию и созданию различных подсистем, таких как системы управления, безопасности или пожаротушения. Плотность размещения оборудования в модулях, как правило, выше, чем в обычных центрах, и они занимают меньше места, а производительность их зачастую оказывается выше, поскольку недостатки конструкции выявляются еще на этапе проектирования модулей.

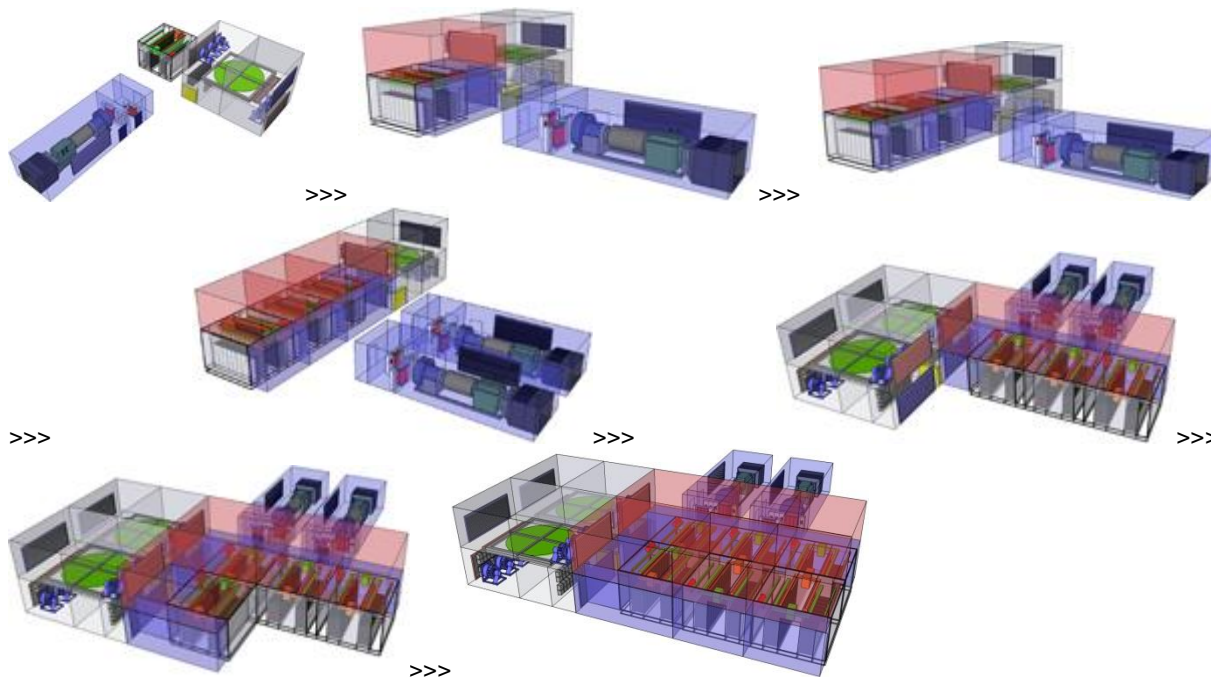


Рис. 6. Концепция создания модульного ЦОД (источник Stack.Net).

МЦОД могут быть размещены практически на любой площадке, в том числе, на территориях, где капитальное строительство затруднено или невозможно. Последнее обстоятельство важно ещё и потому, что МЦОД может быть размещён в непосредственной близости к месту его использования, что означает возможность значительного увеличения скорости обмена данными между серверами и системами хранения данных, размещённых в ЦОД, и вычислительными системами пользователей его услуг.

Поскольку каждый из модулей производится и тестируется в заводских условиях, МЦОД более надёжны и более просты в эксплуатации. Известно, что многие проблемы, которые возникают при эксплуатации ЦОД, являются результатом погрешностей, допущенных во время их конструирования, таких, как расчётные ошибки проектирования (недостаточные энергетические мощности, неэффективное охлаждение и т.п.), в выборе устройств, ошибки, связанные с неправильным размещением или подсоединением устройств.

В связи с возрастающей стоимостью энергии и с тем, что центры обработки данных расходуют все большее количество энергии для удовлетворения возрастающих потребностей клиентов, энергоэффективность ЦОД становится одной из его ключевых характеристик. Здесь МЦОД также имеют важные преимущества перед ЦОД, они заранее оптимизированы в отношении потребления энергии и эффективности охлаждения [14].

И, наконец, одно из главных преимуществ МЦОД – пользователь имеет возможность использовать (и оплачивать) именно тот объём вычислительных

мощностей, который ему необходим, и возможность относительно легко и недорого нарастить или сократить ресурсы за счёт увеличения или уменьшения количества модулей. Кроме того, уменьшается негативный эффект от быстрого морального устаревания технического оснащения ЦОД. По самым оптимистическим расчетам полностью сформированный центр способен быть актуальным на рынке услуг всего 3 года, при этом не учитывается время, уходящее на постройку здания, оснащение и настройку оборудования. МЦОД обладает существенно более простыми, быстрыми и менее затратными возможностями модернизации основных элементов серверного, телекоммуникационного и инженерного оборудования.

Таким образом, использование МЦОД позволяет относительно недорого и в разумные сроки получить в своё распоряжение необходимые вычислительные мощности, имея при этом возможность по необходимости изменить их объём в большую или меньшую сторону. Всё это делает именно МЦОД идеальной инфраструктурой для развёртывания облачных вычислений, особенно в модели частного облака [15], что представляется наиболее разумным вариантом предоставления необходимых вычислительных мощностей для нужд современной вычислительной биологии.

ЗАКЛЮЧЕНИЕ

В условиях значительного увеличения количества и скорости прироста экспериментальных данных, необходимости решения задачи по оперативной обработке этих данных, характерных для многих областей современной науки, в частности для вычислительной биологии и биоинформатики, использование технологий облачных вычислений представляется оптимальной возможностью получения необходимых вычислительных мощностей. При этом использование именно модульных центров обработки данных является наиболее рациональным решением для создания облачных вычислительных структур в силу их дешевизны, простоты и оперативности их развёртывания и ввода в эксплуатацию, что особенно актуально для научных центров в современных реалиях.

СПИСОК ЛИТЕРАТУРЫ

1. Report to the President. Computational Science: Ensuring America's Competitiveness. President's Information Technology Advisory Committee. June 2005. *Networking and Information Technology Research and Development (NITRD) Program*. URL: http://www.nitrd.gov/pitac/reports/20050609_computational/computational.pdf (дата обращения: 25.11.14).
2. Хаубольд Б, Вие Т. *Введение в вычислительную биологию. Эволюционный подход*. Ижевск: НИЦ «Регулярная и хаотическая динамика», Ижевский институт компьютерных исследований, 2011.
3. Леск А.М. *Введение в биоинформатику*. М.: БИНОМ. Лаборатория знаний, 2009.
4. Cycle Computing Website. URL: <http://www.cyclecomputing.com/products-solutions/cyclecloud/> (дата обращения: 25.11.14).
5. Исаев Е.А., Корнилов В.В., Тарасов П.А. Научные компьютерные сети – проблемы и успехи в организации обмена большими объемами научных данных. *Математическая биология и биоинформатика*. 2013. Т. 8. № 1. С. 161–181. doi: [10.17537/2013.8.161](https://doi.org/10.17537/2013.8.161)
6. Исаев Е.А., Корнилов В.В. Проблема обработки и хранения больших объемов научных данных и подходы к ее решению. *Математическая биология и биоинформатика*. 2013. Т. 8. № 1. С. 49–65. doi: [10.17537/2013.8.49](https://doi.org/10.17537/2013.8.49)
7. Amazon cloud HPC hits a “petaflop”. *Ars Technica Addendum*. URL: <http://arstechnica.com/information-technology/2013/11/18-hours-33k-and-156314-cores-amazon-cloud-hpc-hits-a-petaflop/> (дата обращения: 25.11.14).

8. *TOP500 Supercomputer Sites*. TOP500.org. URL: <http://www.top500.org/lists/2014/11/> (дата обращения: 25.11.14).
9. *2013 Cost of Data Center Outages*. Ponemon Institute. URL: http://www.emersonnetworkpower.com/documents/en-us/brands/liebert/documents/white%20papers/2013_emerson_data_center_cost_downtime_sl-24680.pdf (дата обращения: 25.11.14).
10. *Tier Standard: Operational Sustainability*. The Uptime Institute. 2013.
11. *Global Mega Data Center Market 2014-2018*. Infiniti Research Limited (Technavio). Published: Nov 5, 2014. URL: <http://www.giiresearch.com/report/inf273295-global-data-center-market.html> (дата обращения: 25.11.14).
12. *Worldwide Datacenter Census and Construction 2014–2018 Forecast: Aging Enterprise Datacenters and the Accelerating Service Provider Buildout*. International Data Corporation. Oct 2014. Doc # 251830. URL: <http://www.idc.com/> (дата обращения: 25.11.14).
13. Российский рынок коммерческих дата-центров 2013-2017. *iKS-Consulting*. URL: <http://www.iksconsulting.ru/reports-57.html> (дата обращения: 25.11.14).
14. *The Green Grid Data Center Power Efficiency Metrics: PUE and DCiE*. The Green Grid Consortium. URL: <http://www.thegreengrid.org/Global/Content/white-papers/The-Green-Grid-Data-Center-Power-Efficiency-Metrics-PUE-and-DCiE> (дата обращения: 25.11.14).
15. Оплачко Е.С., Устинин Д.М., Устинин М.Н. Облачные технологии и их применение в задачах вычислительной биологии. *Математическая биология и биоинформатика*. 2013. Т. 8. № 2. С. 449–466. doi: [10.17537/2013.8.449](https://doi.org/10.17537/2013.8.449).

Материал поступил в редакцию 27.01.2015, опубликован 12.02.2015.