

Параллельный доступ к данным в HPC-средах: подход SGI

Публикация продолжает серию материалов, посвященных специализированным решениям для высокопроизводительного доступа к данным (SN № 4/33, 2007), представленным на региональном рынке.

Введение

В последние годы наблюдается рост интереса к высокопроизводительным вычислениям как в мировом масштабе, так и в России. В списке самых высокопроизводительных суперкомпьютеров мира — top500.org, благодаря прошлогодним инсталляциям, Россия делит сейчас 10-е место по количеству установленных систем, что можно считать безусловным достижением. Развитие промышленности, нефтегазовой отрасли, государственного сектора, научных и образовательных учреждений ставит задачи построения высокопроизводительных компьютерных систем (ВКС), позволяющих существенно ускорить процессы производства и улучшить качество выпускаемой продукции. Однако из-за относительной молодости этого направления на российском рынке, у заказчиков возникает ряд проблем, связанных с оптимизацией и управлением существующих и создаваемых ВКС.

Проектирование и внедрение высокопроизводительных вычислительных систем

Современное развитие производства немислимо без использования новых компьютерных технологий. И чем сложнее конечный продукт, услуга или процесс, чем сложнее виртуальная модель изделия — тем мощнее и производительнее должна быть вычислительная система, чтобы решать поставленные задачи. Мощностей одиночных стандартных серверов сегодня зачастую недостаточно для решения возникающих задач, таким образом, предприятиям приходится задумываться о внедрении высокопроизводительных вычислительных систем для обеспечения конкурентоспособности. Появление современных многоядерных процессоров вкупе с уменьшением стоимости компьютерных систем ведет к тому, что проблема проектирования и внедрения высокопроизводительных вычислительных систем сегодня как никогда остра.

Еще несколько лет назад термин “высокопроизводительные вычисления” в основном относился к большим многопроцессорным серверам с большим объемом общей памяти. Однако возросшая

производительность современных процессоров, способность устанавливать в стандартные серверы достаточно большой (до 64 Гбайт) объем памяти, всемерное развитие операционной системы Linux при снижении стоимости привели к тому, что сегодня доминирующее положение на рынке высокопроизводительных систем занимают кластеры.

SN в предыдущих номерах рассматривал обзор существующих архитектур для высокопроизводительных вычислений, поэтому подробно останавливаться на этом мы не будем.

В настоящий момент оптимальным решением по показателю цена/производительность для высокопроизводительных систем являются кластеры. Все ведущие вендоры сегодня предлагают кластерные решения. Кластеры в общем случае — это система связанных вычислительных стандартных компьютеров (узлов), соединенных высокоскоростными (Infiniband, Gigabit Ethernet и т.д.) интерфейсами, на которых выполняются специально написанные с помощью параллельных библиотек (в основном MPI) приложения. Большой частью это кластеры на процессорах Intel, однако решения на процессорах AMD и IBM тоже имеют существенную долю рынка. В основном кластеры работают под управлением ОС Linux, однако в последнее время используются и другие платформы, например Microsoft Windows Compute Cluster Server (CCS). Практически все существующие сегодня задачи, будь то инженерные и научные расчеты, обработка сейсмички, изобретение новых лекарственных и химических препаратов, математическое моделирование и исследования физических процессов способны эффективно решаться на кластерах. Более 80% из 500 самых производительных компьютерных систем по списку top500.org — это кластеры.

Однако существует целый ряд задач, которые не очень эффективно работают на кластерах из-за сложности программирования или из-за требований к большому объему оперативной памяти и ресурсам ввода-вывода. Для таких задач оптимальным является использование многопроцессорных систем с общей памятью с

большим количеством процессоров и большим объемом оперативной памяти. К таким задачам могут относиться сложные космические, метеорологические исследования и предсказания погоды, очень сложные научные и инженерные расчеты, а также некоторые унаследованные приложения, которые нет необходимости переносить на новые платформы. Кроме того, такие системы очень эффективно использовать для высокопроизводительных бизнес-приложений и баз данных.

Однако редко где установленная высокопроизводительная система используется для решения только одной задачи, в основном таких задач много и, возможно, лучшим решением будет использование гибридной архитектуры: кластер + суперкомпьютер. Но тем не менее, еще раз хочется подчеркнуть — в общем случае выбор архитектуры высокопроизводительной системы зависит исключительно от задач, которые предстоит решать заказчику.

Помимо задачи выбора собственно вычислительной системы, перед заказчиком стоит задача ввода в эксплуатацию, эффективного использования и управления вычислительной системы. Это также важно, так как, например, ранее процесс ввода в эксплуатацию занимал существенное время (до нескольких месяцев).

Не менее важной задачей является проблема обеспечения инженерной инфраструктуры Центра обработки данных. В современных высокопроизводительных решениях вопрос обеспечения эффективности электропитания и системы охлаждения играет существенную роль, так как мы говорим о десятках и сотнях вычислительных узлов, что требует пересмотра подхода к проектированию серверных помещений.

Ну и, наконец, чрезвычайно важной задачей является выбор наиболее подходящей для решения конкретных задач системы хранения, позволяющей наиболее эффективно обеспечивать данными вычислительную систему. Причем, здесь речь идет не только о выборе, например, дискового массива, но и о системе хранения данных в целом, так как в кластерных средах необходимо обеспечить параллельный доступ к данным десят-

кам, сотням и даже тысячам вычислительных узлов.

Таким образом, говоря о высокопроизводительных вычислительных системах, необходимо понимать, что в действительности заказчик должен решать целый комплекс связанных проблем, начиная от проектирования помещения и заканчивая системами хранения данных. Иначе говоря, вендоры и интеграторы должны предлагать заказчику единый комплекс решения для высокопроизводительных вычислений.

Архитектура SGI HPC-систем

Компания SGI уже более 25 лет фокусирует свою деятельность на рынок высокопроизводительных вычислений, где ей удалось достичь значительных успехов благодаря технологическим достижениям компании как в вычислительной области, так и в области систем хранения данных. По общему количеству установленных систем в списке top500.org компания занимает 4-е место, 10 из 100 самых производительных систем – производства SGI, в том числе система SGI Altix ICE, установленная в NMCAC – The New Mexico Computing and Applications Center, занимающая сейчас 3-е место в списке top500.org и имеющая 14336 процессорных ядер и 28,7 Тбайт памяти, а также система SGI Altix 4700, установленная в LRZ (Leibniz Rechenzentrum в Мюнхене), имеющая 9728 процессорных ядер и 38 Тбайт глобальной общей памяти. Компания SGI много сделала также для всего Linux-сообщества, отдав свои ключевые разработки в открытый код Linux.

В области вычислительных систем SGI предлагает сегодня на рынок решения на базе двух архитектур: кластерные системы на базе процессоров Intel Xeon и суперкомпьютерные системы глобальной общей памяти на базе процессоров Intel Itanium2 и архитектуры SGI NUMAflex.

Кластерные системы SGI Altix ICE (Integrated Compute Environment), построенные на блэйд-архитектуре, имеют высочайшую плотность вычислений (до 64 узлов – 512 процессорных ядер в одном шкафу), интегрированные 2 канала Infiniband, что дает возможность строить кластерные системы, состоящие из тысяч узлов.

Системы SGI Altix на базе архитектуры с глобальной общей памяти масштабируются до 1024 процессоров и 128 Тбайт глобальной общей памяти под управлением единого образа ОС Linux. Они также построены на основе блэйд-архитектуры, что позволяет обеспечить высокую плотность вычислений и легкость независимого масштабирования системы.

Используя свой многолетний опыт, компания SGI в своих системах использует эффективную систему электропитания, а также предусматривает возможность водяного охлаждения вычислительных систем, что является важным фактором построения современных центров обработки данных.

Важнейшим фактором является также тот факт, что компания SGI предлагает полностью интегрированные на фабрике кластерные решения с предустановленным набором системного программного

обеспечения. Этот набор включает: предустановленную операционную систему Linux Suse SLES или Redhat RHEL, SGI ProPack – надстройку над стандартным Linux, куда входят усовершенствования, разработанные SGI для повышения эффективности высокопроизводительных вычислений, программный продукт управления кластером Scali Manage, программное обеспечение создания и управления очередями заданий Altair PBS Pro, а также набор библиотек и средств разработки для параллельных вычислений. Таким образом, компания предлагает комплексный продукт, максимально готовый к промышленной эксплуатации.

Методология построения решений для параллельного доступа к данным

Высокопроизводительный параллельный доступ к данным всех узлов высокопроизводительных систем, зачастую состоящих из тысяч процессоров и узлов кластера, является одним из важнейших факторов обеспечения эффективности их работы.

Современные методы обеспечения параллельного доступа к данным основываются на двух основных подходах: традиционные NAS/NFS решения и решения на основе нового объектного подхода доступа к файлам. В последнее время традиционные NAS/NFS решения все более совершенствуются, улучшая свою производительность и масштабируемость, используя, например, преимущества кластеризации серверов NAS, однако мы, по традиции, будем говорить о них как о традиционных NAS решениях, имея в виду NAS/NFS решения последующих поколений, основанные на новых стандартах. Оба этих решения имеют свою нишу на рынке, таким образом, заказчики имеют возможность выбора в зависимости от требований задач и бюджета сегодняшнего дня, а также возможности масштабирования в будущем.

На рынке существует сегодня большой выбор NAS решений, предлагаемых крупнейшими вендорами, такими как NetApp, Exanet, Isilon, HDS/BlueArc, HP, EMC и др. Компания SGI предлагает на выбор несколько вариантов решения проблемы параллельного высокоскоростного доступа к данным. Основным критерием выбора, по мнению компании, может являться максимальная требуемая производительность параллельной файловой системы. В зависимости от этого значения (1,5, 3, 6, 10 Гбайт/с и более), а также в зависимости от конкретных требований и бюджета заказчиков, компания предлагает различные пути решения из своего широкого спектра предложений по параллельным системам хранения.

НПС системы хранения для небольших кластерных систем с производительностью до 1,5 Гбайт/с

В качестве NAS (Network Attached Storage) решения компания SGI предлагает на рынок линейку NAS устройств SGI Nexis, основанную на высокопроизводительных и масштабируемых во всех направлениях серверах SGI Altix, дисковых

массивах SGI InfiniteStorage и программного пакета SGI Appliance Manager. В основе решения лежит локальная файловая система SGI XFS. Эта 64-битная файловая система с журналированием была разработана компанией SGI еще в 1995 г., она обладает высочайшей производительностью и способна масштабироваться до 18 млн терабайт. Следуя своей стратегии ориентации на открытые системы, компания SGI еще в 2001 г. отдала исходный код своей файловой системы XFS сообществу Linux, и, начиная с ядра 2.4, файловая система XFS входит в код ядра Linux.

Кроме того, SGI оптимизировала стек NFS протоколов, что позволило достичь производительности стандартного протокола NFS до 1,5 Гбайт/с. Таким образом, в небольших кластерах, где такой производительности достаточно, SGI предлагает использовать NAS решения SGI Nexis.

НПС системы хранения для “средних” кластерных систем с производительностью до 3 Гбайт/с

К сожалению, стандартный протокол NFS в сегодняшней имплементации практически ограничен скоростями в 1,5 Гбайт/с. Многие вендоры предлагают сегодня так называемые кластерные NAS-решения, способные преодолеть этот барьер и добиться скоростей до 10 Гбайт/с. Однако широкое внедрение новых высокоскоростных кластерных интерконнектов (например Infiniband 4X DDR со скоростью 20 Гбит/с) привело к мысли использовать их не только как вычислительную сеть, но и как среду для высокоскоростного доступа к данным. Появились дисковые массивы (SGI предлагает их в том числе), использующие Infiniband в качестве хост-соединений, заменяя Fibre Channel (или давая возможность совместного использования FC- и IB-соединений). Таким образом, появилась возможность прямого доступа к высокопроизводительным массивам с вычислительных узлов кластера, в том случае, если им нужен быстрый доступ, например, для scratch данных. Однако это не решало проблему параллельного совместного высокоскоростного доступа к данным. Возможности протокола Infiniband, например, RDMA (Remote Direct Memory Access), привело к мысли использовать их для ускорения доступа к данным. Компания SGI стала первым вендором, которая представила на рынок решение на базе протокола NFS-over-RDMA. Старшая модель NAS-решений – SGI Nexis 9000 использует “родной” протокол Infiniband и позволяет добиться производительности NFS до 5 Гбайт/с на чтение и до 3 Гбайт/с на запись. Возможность использования существующей вычислительной Infiniband кластерной инфраструктуры позволяет использовать единую сеть как для вычислительных ресурсов, так и для доступа к данным, что ведет к существенной экономии средств. А в случае использования нескольких каналов Infiniband (как в случае, например блэйд-решений SGI Altix ICE), имеется возможность как раздельного использования высокоскоростных каналов (вычислительный и доступа к данным), так и использования их в отказоустойчивой среде.

Таким образом, для удовлетворения требований скорости параллельного доступа до 3 Гбайт/с в кластерной среде, SGI предлагает решения, основанные на открытых стандартах, легко внедряемые и не требующие модификаций клиентских составляющих.

НПС-системы хранения для больших кластерных систем с производительностью свыше 50 Гбайт/с

Однако зачастую требования параллельного доступа к данным в больших кластерных системах существенно превышают 3 Гбайт/с. В этом случае предпочтительнее использовать параллельные файловые системы с объектной архитектурой, такие, например, как PanFS от компании Panasas или разрабатываемая по открытой лицензии GNU GPL файловая система Lustre от компании Cluster File Systems, купленной в конце 2007 г. компанией Sun Microsystems, а также рассмотренная в предыдущих номерах SN файловая система GPFS от компании IBM.

НПС системы хранения с объектной архитектурой

Если коротко, то парадигма объектной архитектуры заключается в том, что каждый файл ассоциируется со множеством объектов, находящихся на разных физических устройствах. При выполнении файловых операций вычислительный узел получает карту с указанием расположения объектов на физических устройствах и далее имеет возможность напрямую работать с объектами, хранящимися на них, без использования выделенного сервера. Таким образом, обеспечивается множество параллельных путей доступа к данным с вычислительных узлов, что и позволяет добиться великолепной масштабируемости системы хранения и обеспечения высочайшей производительности (до 50 Гбайт/с и более).

Компания SGI предлагает в качестве решений с объектной архитектурой PanFS от компании Panasas и Lustre от компании Cluster File Systems (точнее — уже от компании Sun Microsystems).

Решение от компании Panasas представляет собой комплексное программно-аппаратное решение на основе блэйд-архитектуры, которое легко развертывается и масштабируется. SGI предлагает это решение в том случае, если требуемая производительность системы хранения находится в диапазоне от 3 до 6 Гбайт/с.

В том случае, если требуемая производительность превышает 6 Гбайт/с, компания SGI предлагает использовать параллельную файловую систему Lustre на дисковых массивах SGI InfiniteStorage, способных обеспечивать одинаковую производительность как чтения, так и записи до 6 Гбайт/с и обладающих высочайшей плотностью хранения (600 дисков в одном шкафу) и масштабируемостью (1200 дисков x 1 Тбайт = 1,2 Пбайт). Файловая система Lustre позволяет использовать стандартное аппаратное обеспечение и поддерживает широкий спектр интерконнектов, не ограничиваясь Gigabit Ethernet, что позволяет системе масштабироваться до тысяч узлов и до сотни гигабайт в секунду.

Использование параллельной файловой системы Lustre совместно с дисковыми массивами SGI InfiniteStorage позволяет добиваться производительности свыше 50 Гбайт/с и масштабироваться до тысяч узлов. Поскольку проектирование и начальная установка такой системы достаточно сложны, эти операции проводятся профессионалами из подразделения SGI Professional Services, таким образом, заказчик получает комплексное решение, устанавливаемое и поддерживаемое компанией SGI.

В частности, система SGI Altix ICE, установленная в NMCAC — The New Mexico Computing and Applications Center — и занимающая сейчас 3-е место в списке top500.org и имеющая 14336 процессорных ядра и 28,7 Тбайт памяти, обеспечивает параллельный доступ с помощью 172ТБ файловой системы Lustre.

НПС системы хранения на основе SAN-инфраструктуры

Однако, помимо рассмотренных выше решений параллельного доступа, основанных на базе NFS/NAS архитектуры и объектной архитектуры, существуют общие файловые системы, работающие в инфраструктуре SAN (Storage Area Network). Такими файловыми системами являются, например, Quantum/Adic StorNext и SGI CXFS.

SGI CXFS, основанная на файловой системе SGI XFS, представлена на рынке с 1999 г. и является первой общей файловой системой SAN в индустрии.

Архитектура SAN при всех своих преимуществах имеет один недостаток — отсутствие совместного доступа к данным, который в общем случае приходится решать традиционными способами: разделением массива на несколько частей, соответствующих различным операционным системам, и обеспечивая общий доступ через локальную сеть. Компания SGI устраняет этот недостаток, представляя гетерогенную общую файловую систему SGI CXFS. Эта отлично зарекомендовавшая себя временем высокоскоростная 64-битная файловая система с журналированием, позволяет клиентам под Windows, Linux, IRIX, Solaris, AIX и MacOSX работать с данными и теми же данными со скоростями до 45 Гбайт/с. Использование общей файловой системы SGI CXFS позволяет по-новому взглянуть на технологический процесс, обеспечивая совместное высокоскоростное использование централизованного хранилища данных на всех его этапах. Но что еще более важно, это решение, основанное на открытых технологиях, позволяет использовать в архитектуре SAN прикладные решения от многих производителей программного и аппаратного обеспечения. Кроме того, нельзя не упомянуть о еще одной уникальной особенности SGI CXFS — гарантированной скорости ввода-вывода. То есть любой системе, участвующей в SGI CXFS, можно выделить полосу пропускания и быть уверенными, что система будет работать с данными с указанной скоростью. Это особенно важно в медиа-индустрии, где работа с гарантированными потоками в реальном времени зачастую является производственной необходимостью.

SGI CXFS широко используется также как средство высокоскоростного общего доступа к данным в кластерных системах, где вычислительные узлы представляют собой так называемые “толстые узлы”. В основном, это серверы SGI Altix, с числом процессоров на узел до 512-1024. Например, система SGI Altix, установленная в исследовательском центре NASA и на момент установки являвшаяся самой производительной вычислительной системой в мире, состоит из 20 512-процессорных систем. Используя высокоскоростную 440ТБ SAN систему хранения под управлением общей файловой системы SGI CXFS, “толстые” узлы SGI Altix получают высокоскоростной общий доступ к SAN-хранилищу, причем каждый узел может работать со скоростями Fibre Channel, то есть со скоростями до десятков гигабайт в секунду.

Клиент-серверная архитектура SGI CXFS подразумевает использование нескольких серверов метаданных (MDS) для отказоустойчивости и балансировки нагрузки. Современные архитектуры SAN также позволяют создавать полностью отказоустойчивые сети хранения (например, с использованием двух независимых фабрик), что вкупе с использованием избыточных каналов Fibre Channel в серверных/клиентских системах, функцией автоматического обхода неисправностей, реализованной в менеджере логических дисков SGI XVM, дает возможность строить полностью отказоустойчивые системы с отсутствием единой точки отказа.

Особую популярность SGI CXFS получила в индустрии кинопроизводства. Практически все кинофильмы снимаются на пленку, однако далее исходный материал переводится в цифровой вид и вся дальнейшая работа ведется с файлами. Далее идут процессы цветокоррекции, монтажа, компьютерной графики и т.п. И только финальный результат переводится опять на пленку и показывается в кинотеатрах. Весь этот процесс носит название Digital Intermediate (DI). Все хорошо, но дело в том, что отсканированный в формате 2K (2048x1556) кадр имеет размер 12,5 Мбайт, 24 кадра в секунду дают нам поток в 300 Мбайт/с, исходные материалы могут занимать несколько терабайт! Понятно, что при отсутствии высокоскоростной общей файловой системы в технологическом процессе DI, участникам процесса (колористам, монтажерам, художникам) пришлось бы копировать материал (а это терабайты!) к себе на локальный компьютер, производить там изменения и копировать обратно! То есть значительную часть рабочего времени проводить в ожидании, пока материал скопируется. Сложно говорить в этом случае о полноценной коллективной работе, не говоря о том, что многотерабайтные копии фильма способны заполнить все локальные диски. И, конечно, не говоря уже об обеспечении проигрывания 300 Мбайт/с в реальном времени (а режиссеры не любят ждать). А что будет, когда киноиндустрия перейдет на формат 4K (4096x3112), что в 4 раза больше и один поток будет составлять 1,2 Гбайт/с?

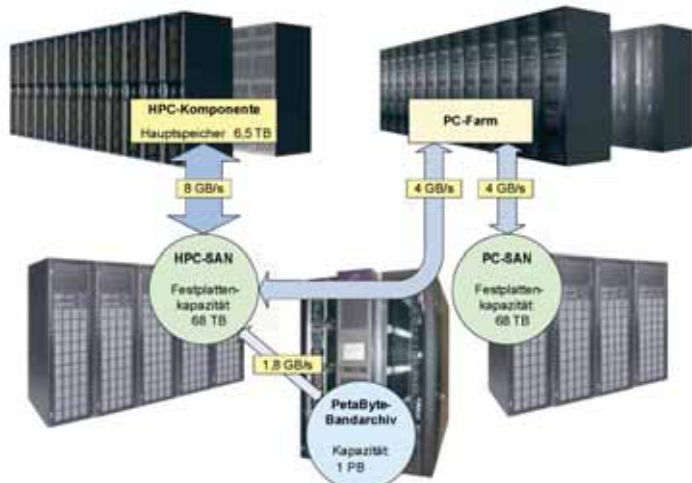


Рис. 1. HPC-проект в Дрезденском техническом университете объединил SGI SMP-систему (Altix 4700) с общим объемом памяти 6 Тбайт, HPC-кластер (2592 ядер), HPC-SAN и PC-SAN – все на базе SGI InfiniteStorage 6700 – DDN S2A9500, а также вторичную память на основе Sun STK8500.

Практически все крупнейшие голливудские студии используют SGI CXFS в качестве решения высокоскоростного общего доступа к файлам. В последние годы мы наблюдаем огромный рост киноиндустрии и в России. Крупнейшие российские студии также используют решения SGI CXFS в своей работе.

Например, молодая, но быстро растущая группа компаний “Синелаб” недавно расширила свой спектр деятельности (помимо услуг аренды кинокамер, света и другого оборудования для кинопроизводства, а также современной лаборатории) открытием постпродакшн студии. Предлагая полный спектр услуг Postproduction и Digital Intermediate, компания Cinelab Digital располагает современнейшими системами цифровой цветокоррекции Autodesk Lustre Incinerator, монтажными системами Autodesk Smoke, Avid Media Composer Adrenaline и FinalCut Pro, устройствами ввода-вывода изображения с/на пленку, такими как ARRILASER/ARRISCANNER, Thomson Spirt, Pandora. Для обеспечения высокоскоростного общего доступа всех этих компонентов к материалу, компания “Синелаб” выбрала комплексное решение SGI SAN/CXFS. Оперативная система хранения построена с использованием дискового массива SGI InfiniteStorage 6700, способного обеспечить работу одновременно с 6-ю потоками 2К (более 1800 Мбайт/с) в реальном времени. Вторичная система хранения построена с использованием дискового массива SGI InfiniteStorage 350 на SATA-дисках, что позволяет обеспечивать экономичное временное хранение материалов. Для долговременного хранения используется ленточная библиотека Quantum/Adic i500.

Использование общей файловой системы SGI CXFS позволяет всем участникам технологического процесса компании обеспечивать высокоскоростной общий доступ к материалу с разных операционных систем в реальном времени, не тратя время и место на копирование на локальные массивы, что позволяет существенно ускорить технологический процесс и предлагать быстрый и высококачественный сервис компаниям-заказчикам. Используемая архитектура SGI позволяет легко масштабировать систему при переходе от работы с 2К

(300МБ/с) до 4К (1200МБ/с) потоками.

Интересно также было бы рассказать о комплексном примере внедрения решения SGI в Техническом Университете Дрездена (Technische Universität Dresden). SGI как генподрядчик проекта внедрил здесь практически весь комплекс предлагаемых компанией продуктов и решений (рис. 1, http://tu-dresden.de/die_tu_dresden/zentrale_einrichtungen/zjh/lehre/ws0708/lars/vorlesung_1.pdf).

В составе вычислительной части проекта были представлены как SGI Altix с 2048 процессорными ядрами Intel Itanium2 и 6,5 Тбайт глобальной общей памяти, так и кластером из 728 вычислительных узлов. В качестве системы хранения использовались массивы SGI InfiniteStorage 6700. Часть системы хранения работает под управлением общей файловой системы SGI CXFS, обеспечивая доступ серверам SGI Altix со скоростями до 8Гбайт/с. В то же время, серверы CXFS обеспечивают SAN/NAS конвергенцию, предоставляя кластеру NFS доступ со скоростью до 4 Гбайт/с к SAN/CXFS хранилищу. Иерархическая система хранения и управления жизненным циклом данных SGI DMF (Data Migration Facility) обеспечивает прозрачную миграцию малоиспользуемых данных на ленточную библиотеку Sun StorageTek 8500 со скоростью до 1,8 Гбайт/с. Другая часть системы хранения работает под управлением параллельной файловой системы Lustre, обеспечивая высокоскоростной доступ к хранилищу со скоростями до 4 Гбайт/с. Таким образом, мы видим, что даже в рамках одного проекта могут использоваться различные схемы организации параллельного доступа в зависимости от требований задач.

Вместо заключения

Таким образом, мы видим, что процесс построения высокопроизводительных систем является достаточно сложным, требующим предусмотреть множество нюансов, в конечном итоге влияющих на комплексную производительность системы в целом. Мы рассмотрели комплексный подход компании SGI к построению высокопроизводительных систем. Однако, как уже упоминалось, практически все крупнейшие вендоры и интеграторы предлагают свои пути решения возникающих проблем, в том числе в области параллельного доступа к данным. Индустрия сегодня ждет появления стандарта pNFS (NFS 4.1), параллельного NFS, который, надеемся, позволит стандартизировать существующие сегодня решения и вывести решения параллельного доступа к данным на новый уровень.

Александр Анциферов, группа компаний “Арбайт”, технический директор по продуктам SGI

Оптическая технология для высокоскоростного обмена файлами

Февраль 2008 г. – Исследователи корпорации IBM представили прототип технологии, которая позволит создавать коммуникационные каналы с высокой пропускной способностью и низким энергопотреблением для всех разновидностей компьютеризированных систем – от суперкомпьютеров до сотовых телефонов.

Новая технология, которая использует для передачи информации волноводы вместо проводов, способна обеспечивать обмен данными на скоростях до 8 Тбит/с, потребляя при этом количество электроэнергии не более, чем одна 100 Вт лампочка.

Уже создан набор микросхем для оптического трансивера и сам трансивер – приемопередатчик с большим числом параллельных каналов и повышенным быстродействием, объединяющий 24 передатчика и 24 приемника, каждый из которых работает со скоростью 12,5 Гбит/с. Т.о., суммарная двунаправленная скорость обмена данными достигает значения в 300 Гбит/с, что вдвое превышает быстродействие устройств предыдущего поколения. В сравнении с существующими коммерчески доступными сетевыми оптическими модулями, новый трансивер демонстрирует десятикратно большую пропускную способность при десятикратно меньших габаритах и сопоставимом энергопотреблении. Для поддержки возможности экономически выгодного массового производства новый трансивер использует стандартные 850-нанометровые плоскостные лазеры с вертикальным резонатором (850-nm vertical-cavity surface emitting laser, VCSEL), представляющие собой высокоскоростную версию недорогой технологии, которая применяется в компьютерных оптических мышах.

Суть новой технологии в том, что в печатных платах с оптическими электронными элементами – или “Optocards”, “оптоэлектронных платах” – в качестве электрическими проводниками между передатчиками и приемниками сигналов применяется массив полимерных оптических волноводов с малыми потерями. Шина данных, построенная на основе таких оптоэлектронных плат, не только формирует большое число высокоскоростных внутренних коммуникационных каналов, но также “тесно” объединяет их на плате, позволяя достичь беспрецедентно высокой плотности монтажа – каждый отдельный волноводный канал тоньше человеческого волоса. Высокая плотность монтажа очень важна для электронного устройства, поскольку допускает интеграцию гибридных элементов в микросхеме, что, в свою очередь, позволяет создавать высокоинтегрированные оптоэлектронные модули или “опточипы” (“Optochip”).