

КЛАСТЕРНАЯ ВЫЧИСЛИТЕЛЬНАЯ СИСТЕМА ЦЕНТРА ОБРАБОТКИ ДАННЫХ ВЯТСКОГО ГОСУДАРСТВЕННОГО УНИВЕРСИТЕТА

В.С. Князьков, А.В. Частиков., А.Н. Краев

Кластерная вычислительная система центра обработки данных Вятского государственного университета спроектирована как система с распределенной памятью и построена на базе 288 однотипных вычислительных узлов универсального назначения, объединенных двумя специализированными коммуникационными средами. Использование независимых коммуникационных сред позволило исключить взаимное влияние потоков управления и потоков данных и повысить скорость исполнения прикладных программ.

Доступ к пользовательским приложениям и вычислительным ресурсам системы, административное управление вычислениями обеспечивается средствами выделенного управляющего сервера.

ТЕХНИЧЕСКИЕ ХАРАКТЕРИСТИКИ БАЗОВОГО ОБОРУДОВАНИЯ СИСТЕМЫ

Количество вычислительных узлов системы – 288. Общее количество вычислительных ядер – 1152. Количество выделенных управляющих узлов - 2. Суммарный объем оперативной памяти- 584GB. Суммарный объем дисковой памяти на вычислительных узлах – 10 368 GB. Суммарный объем дисковой памяти на управляющих узлах - 288 GB.

Эксплуатационные показатели производительности системы, рассчитанные на основе оценки результатов тестирования производителем сегмента кластера из 9-ти вычислительных узлов на тесте HPL (High Performance Linpack), следующие.



1. Пиковая производительность – 19 000,00 GFlops (миллиардов операций с плавающей точкой в секунду).
2. Соотношения R_{max}/R_{peak} составляет не менее 65%, R_{max} – оценка максимальной производительности системы по результатам тестирования сегмента кластера из 9-ти вычислительных узлов тестом HPL, R_{peak} – пиковая производительность системы.
3. Соотношение R_{max}/W_{max} составляет 148,8 GFlops на 1 КВт потребляемой энергии, $W_{max} = 83$ КВт – потребляемая мощности системы при максимальной загрузке,

Вычислительные узлы системы спроектированы на базе блейд-серверов "половинной" высоты на аппаратной платформе HP ProLiant BL460c G1 на платформе 4-х ядерных процессоров Quad-Core Intel Xeon® 5345 EM64T (тактовая частота 2.33 ГГц, процессорный КЭШ - 8 Мб). В результате, каждый вычислительный узел на базе процессоров Quad-Core Intel Xeon® 5345 EM64T содержит восемь полноценных процессоров, частота системной шины - 1333 Front Side Bus, чипсет Intel® 5000P, 2 встроенных однопортовых многофункциональных гигабитных контроллера сетевого интерфейса NC373i и 1 дополнительный контроллер сетевого интерфейса 10/100 для решения iLO 2 Management. Каждый вычислительный узел содержит: оперативную память общим объемом в 4GB - 2-х HP 2GB FBD PC2-5300 (2x1GB) типа PC2-5300 DIMM с полной буферизацией, 667 МГц ; 2 дисковых накопителя типа HP 10K SAS 2.5 Hot Plug Hard Drive с объемом одного дискового накопителя 36GB.

Использованные при построении системы технологические решения поддерживаются современными компиляторами и фактически обеспечивают удвоение эффективной производительности каждого процессора. Низкая стоимость вычислений и возможность последующего масштабирования вычислительного ядра системы обеспечивает системе долгий жизненный цикл.

Управляющий узел построен аналогично вычислительному узлу на платформе четырехядерных процессоров Quad-Core Intel Xeon® 5345 EM64T с тактовой частотой 2.33 ГГц, но имеет размер оперативной памяти на каждое вычислительное ядро узла - 512MB, общий объем дискового пространства - 144GB, а также 2 встроенных сетевых адаптера и реализован в конструктивном исполнении RACK 2U.

Доступ к сети передачи сообщений и сети управления реализован на базе встроенных в каждую серверную полку коммуникационных сетей Infiniband и Gigabit Ethernet.

Вычислительная коммутирующая сеть построена с использованием высокопроизводительной сетевой технологии Infiniband и обеспечивает коммуникации для 288 вычислительных узлов системы. Коммуникационная сеть спроектирована на базе коммутаторов HP BLc Bnt 1GbE2 Switch, HP BLc 4X DDR IB Switch и поддерживает

следующие сетевые интерфейсы - 16 Internal 1Gb Downlinks, 5 External 10/100/1000BASE-T Uplinks, 2 Internal cross-connects, 1 Management Console Port, 16 Internal 4X DDR downlinks, 8 external 4X DDR uplinks, 1 Management Console Port.

Управляющая коммутирующая сеть построена с использованием сетевых технологий Ethernet, Fast Ethernet, Gigabit Ethernet, mini-GBIC на базе коммутаторов HP ProCurve Switch 2848, HP ProCurve 2626 Switch, имеет топологию типа - «Дерево», используемый тип канала - полный дуплекс, ширина канала - 1 Гбит/сек.

Поддерживаются следующие порты и сетевые интерфейсы:

- 44 порта 10/100/1000 (протокол IEEE 802.3, тип 10Base-T, протокол IEEE 802.3u, тип 100Base-TX, протокол IEEE 802.3ab, тип 1000Base-T Gigabit Ethernet);
- 1 консольный порт RS-232C DB-9;
- 4 порта двойного назначения - каждый порт может использоваться как дополнительный порт RJ-45 с автоматическим определением 10/100/1000 или как свободный слот под mini-GBIC (для подключения трансиверов mini-GBIC);
- 24 порта RJ-45 с автоматическим расширением 10/100, протокол IEEE 802.3, тип 10Base-T и протокол 802.3u, тип 100Base-TX;
- 2 порта двойного назначения - каждый порт может использоваться как порт RJ-45 с автоматическим определением 10/100/1000 (протокол IEEE 802.3, тип 10Base-T; протокол 802.3u, тип 100Base-TX; протокол 802.3ab 1000Base-T Gigabit Ethernet) или как свободный слот для mini-GBIC (для подсоединения трансиверов mini-GBIC).

Монтаж оборудования вычислительных узлов и коммуникационных сетей выполнен в шести стандартных 19" 42U шкафах, причем, сохранен визуальный доступ ко всем сигнальным светодиодам и индикаторам на лицевых панелях всех устройств коммуникационной и вычислительной среды.

Система хранения данных построена с использованием решения HP StorageWorks MSA60 – является надежной и простой в управлении системой хранения данных прямого подключения на основе решений SAN. Причем, за счет поддержки жестких дисков SAS и SATA в одном корпусе поддерживается многоуровневая архитектура хранилища данных с отказоустойчивой конфигурацией - для предотвращения простоев или потери данных в период обновления или при сбое одновременно одного или двух дисков реализованы функции RAID-6 и Advanced Data Guarding. Максимальная емкость системы на базе 12 дисков SAS - 3,6 Тбайт, на базе 12 дисков SATA - 9 Тбайт. Предусмотрена также возможность прямого подключения к серверам HP ProLiant и HP Integrity, возможность каскадирования накопителей в конструктиве 8U по схеме 1+3 для установки 48 жестких дисков на каждый из двух портов контроллера Smart Array P800. По мере роста потребности в ресурсах хранения обеспечена возможность подключить до 96 жестких дисков.

Система визуализации системы реализована на базе подсистемы NVIDIA Quadro Plex 1000, смонтированной в 19" стойке. Основными задачами системы визуализации являются:

- сбор информации о системе, ее преобразование в подходящий формат и визуализация; представление информации в виде Web-страниц;
- уведомление о критических событиях в системе (отправка сообщений по указанным адресам электронной почты, на мобильный телефон и т.д.);
- поддержка архива, а также возможность просмотра прошлых состояний кластера (например, за час, день, месяц, год или более).

Кроме этого, система визуализации обеспечивает высокую пропускную способность для пользовательских приложений при визуализации сложных моделей, высокую скорость обработки с высоким разрешением больших текстур и фреймов, обеспечивает синхронизацию внешним источникам данных и исходящими потоками данных (например, при работе с телевещательными приложениями), синхронизацию видеопотоков с нескольких рабочих станций и их мультимплексирование на «единый виртуальный экран».

Спроектированная система визуализации поддерживает форматы видео-вывода 2K, HD и SD, по кадровую синхронизацию в видео потоках. Для поддержки трехмерных эффектов и анимации, обработки полигонных каркасов, ускоренной трассировкой лучей выполняется коррекция изображений для двучетных стерео-дисплеев с использованием нового комплекса Gelato - Gelato 2.1.

СОСТАВ БАЗОВОГО СИСТЕМНОГО ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ

- Операционная система RedHat Enterprise Linux Advanced Server 4 (AMD64-EM64T) .
- HP OpenView Operations - управление коммуникационными сетями, импортированными операционными системами и приложениями, копированием и хранением данных, управление сервисами, в том числе, и Интернет-сервисами.

- HP OpenView Network Node Manager - обеспечивает управление вычислительной сетью. Инструменты, входящие в состав решения HP OpenView NNM, позволяют сократить сроки поиска и устранения технических неисправностей.
- HP OpenView Service Desk – подсистема автоматизации работы служб технической поддержки и управления ИТ-услугами.
- HP TeMIP TeMIP – подсистема управления телекоммуникациями.: Интернет- и локальными сетями, сетями передачи данных. Также обеспечивает управление обработкой отказов и генерацию аварийных сообщений, функции управления и защиты данных — управление паролями, профилями пользователей; функции управления и оптимизации нагрузки сети — функции генерации отчетов о нагрузке, качестве услуг; функции управления конфигурацией аппаратного обеспечения и базы данных, сетевого оборудования и сегментации сетей,
- TelAlert-Vytek – обеспечивает функции оперативного оповещения персонала на стационарные и мобильные телекоммуникационные устройства - прием от абонентов (при двустороннем обмене) управляющих команд и передача команд управления системе; функции групповой рассылки сообщений с учетом индивидуальных и групповых рабочих расписаний, получение подтверждений о доставке рассылаемых сообщений абонентами.
- Report Manager (Westbury) – обеспечивает функции преобразования данных, собираемых на хранение в HP OpenView ServiceDesk в управленческую информацию в виде оперативных, тактических и стратегических отчетов.
- Dashboard Manager - предоставляет возможности для анализа и контроля административной информации.
- Service Request Manager - расширяет возможности Service Desk, позволяя конечным пользователям создавать и отслеживать запрос в службу поддержки, используя web-интерфейс.
- Barcode Manager - представляет собой надстройку к Configuration Management в составе HP OpenView ServiceDesk и позволяет проводить быструю идентификацию и аудит подключенной ИТ-инфраструктуры.
- MS Virtual Server - позволяет разработчикам и администраторам выполнить тестирование и ускорить процесс разработки программного обеспечения, осуществлять миграцию унаследованных бизнес-приложений, обеспечивает возможности восстановления системы в случае сбоя оборудования и кластеризации виртуальных машин на одном или нескольких хостах.
- Xen - монитор виртуальных машин для платформы x86, поддерживает одновременное выполнение нескольких гостевых операционных систем с высокой производительностью и изоляцией ресурсов.
- VMWare - позволяет сформировать виртуальную инфраструктуру на базе нескольких операционных систем.

СОСТАВ ПРИКЛАДНОГО ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ СИСТЕМЫ ВКЛЮЧАЕТ СЛЕДУЮЩИЕ ПРОГРАММНЫЕ КОМПЛЕКСЫ

- CFX - программный комплекс, сочетающий возможности анализа гидрогазодинамических процессов, многофазных потоков, химической кинетики, процессов горения, радиационного теплообмена и многих других. Основные сферы применения: аэрокосмическая, турбомашиностроение, нефтегазовая и химическая промышленность, автомобилестроение и судостроение, теплоэнергетика, системы вентиляции и кондиционирования, биомедицинские приложения, микроэлектромеханические системы.
- Fluent - программный пакет для анализа и расчета химических реакций и процессов горения, сверхзвукового течения газов или жидкостей, расчёта кондиционирования помещений и систем охлаждения электронных схем. Пакет может быть установлен как на PC, так и на высокопроизводительных многопроцессорных вычислительных системах.
- STAR-CD- является многоцелевым CFD-пакетом разработки поддерживающий процедуру “скользящих сеток”. Эффективные механизмы параллелизации вычислений позволяют моделировать задачи высокой геометрической сложности. Области применения STAR-CD: транспорт, аэрокосмическая, автомобильная, железнодорожная, судостроительная промышленность, энергетика - ядерная, гидро, солнечная, топливная (уголь, жидкое топливо, газ), химическая промышленность и общее машиностроение, строительная индустрия, электротехническая и электронная, газовая и нефтяная промышленность, а также биомедицинские приложения.
- LS-DYNA - пакет предназначен для анализа высоконелинейных и быстротекущих процессов в задачах механики твердого и жидкого тела. Основные области применения: нелинейная динамика, теплоперенос, теплообмен, термомеханика, разрушение и развитие трещин, контакт, квазистатика, Эйлера и произвольное Лагранж-Эйлера поведение, акустика, междисциплинарный связанный анализ (взаимодей-

ствие потоков жидкостей и газов с деформируемой конструкцией, например, колебания жидкости в баках, связанные термомеханические задачи и др.)

- ABAQUS - предназначен для проведения прочностного инженерного многодисциплинарного анализа. Основные сферы применения ABAQUS — оборонная промышленность, авиастроение, автомобилестроение, электроника, металлургия, производство энергии, добыча и переработка нефти, производство товаров народного потребления, общая механика и геомеханика. С помощью ABAQUS можно выполнять анализ таких сложных конструкций и механизмов, как турбомашины, двигательные установки, шасси и трансмиссии, шины, транспортные средства, а кроме того решать сложные задачи, такие как сварка, моделирование аварийных столкновений (краш-тесты), тесты на падение и расчет прочности электронных компонентов, сверхпластическое формирование, пробивание материала, расчет композиционных структур, литье металлов, контактное взаимодействие большого числа тел и самоконтакт, сейсмические и взрывные воздействия, расчет надежности ядерных реакторов.
- FlowVision - комплексное решение в области моделирования трехмерных турбулентных течений жидкости и газа. Обеспечивает полный набор расчетов, включая расчет напряженно - деформированных состояний, собственных частот и форм колебаний, анализ устойчивости, решение задач теплопередачи, исследования установившихся и неуставившихся процессов, акустических явлений, нелинейных статических процессов, нелинейных динамических переходных процессов, расчет критических частот и вибраций роторных машин, анализ частотных характеристик при воздействии случайных нагрузок, спектральный анализ и исследование аэроупругости. Предусмотрена возможность моделирования практически всех типов материалов, включая композитные и гиперупругие. Расширенные функции включают технологию суперэлементов (подконструкций), модальный синтез и макроязык DMAP для создания пользовательских приложений.
- MSC/MARC - является универсальной конечно-элементной программой, ее применение особенно эффективно для проведения углубленного анализа высоконелинейного поведения конструкций и решения задач теплопередачи. С помощью MSC/Marc возможно решение задач, по условиям которых конструкции подвергаются большим линейным и угловым перемещениям, материалы имеют нелинейные свойства или свойства, зависящие от истории нагружения, присутствует сложное контактное взаимодействие частей конструкции. Примерами таких задач являются анализ поведения резиновых уплотнений, анализ строительных (в том числе гидротехнических) сооружений с учетом свойств грунта, моделирование контакта зубчатых зацеплений и т.д.
- MAGMASOFT - многофункциональный специализированный мощный CAE-продукт, позволяющий моделировать практические любые виды литейных процессов. MAGMASOFT - позволяет предсказывать дефекты различной природы, рассчитывать структуру сплава, механические свойства, остаточные напряжения и деформации отливок, проводить анализ тепловых и гидродинамических процессов и т.д. Благодаря своим уникальным возможностям MAGMASOFT можно особенно эффективно применять при подготовке производства сложно-профильных и тонкостенных отливок.
- SolidWorks - программный пакет для решения задач: гибридное параметрическое моделирование, проектирование деталей, сборок и изделий с учетом специфики изготовления (листовой материал, пресс-формы и штампы, сварные конструкции), экспресс-анализ (массово-инерционные характеристики, анализ прочности и кинематики), оформление чертежей по ЕСКД.
- Quantum Pharmaceuticals - программный пакет для решения задач молекулярного моделирования для разработки новых лекарств. Технология используется для виртуального скрининга, докинга, расчетов свободной энергии, связывания комплексов белок-лиганд, идентификации и оптимизации кандидатов в лекарственные средства.
- AMBER - программный пакет для решения задач моделирования динамики биологических молекул.
- GROMACS. Пакет программного обеспечения для моделирования динамики биологических молекул.
- GROMOS - программный пакет для решения задач моделирования динамики биологических молекул.
- NAMD - программный пакет для решения задач моделирования динамики биологических молекул.
- BLAST - программный пакет для решения задач сравнения информации о первичных биологических последовательностях, таких как последовательность аминокислот в белках или последовательность нуклеотидов в ДНК. Используя BLAST можно сравнить имеющуюся последовательность с библиотекой или банком последовательностей и определить гомологичные последовательности в библиотеке.
- ASG - программный пакет для решения задач биометрической аутентификации.
- WRF-chem - программный пакет для решения задач моделирования распространения атмосферных примесей и химических реакций.
- Paradigm Geophysical - программный пакет для решения задач интерпретации геолого-геофизических данных.

- Schlumberger - программный пакет для решения задач в области разведки и добычи нефти.
- Landmark – программный пакет для решения задач 3D- интерпретации, визуализации и анализа сейсмических данных.
- Autodesk mental ray - система визуализации, способная осуществлять просчет сцены с использованием Global Illumination, Caustics, Phenomena и Area Lights. Позволяет распределить задание на несколько процессоров или компьютеров в сети и осуществлять параллельную обработку.

Управление оборудованием при Grid-режимах выполнено с использованием возможностей технологии Juniper Networks: поддерживается всесторонний анализ входящего и исходящего трафика на основе 200 статистических параметров.

В процессе вычислений в системе поддерживается балансировка нагрузки на 4-7 уровнях:

- балансировка нагрузки HTTP, HTTPS (SSL), FTP, TCP и UDP трафика;
- полноценная балансировка нагрузки в зависимости от метода запроса, версии протокола, URL, cookie, параметров заголовков и содержимого пакетов, POST-данных, SOAP, XML;
- программируемая проверка работоспособности и функциональности внешних устройств и серверов;
- поддержка усовершенствованного метода балансировки HTTP/S трафика “Fewest Outstanding Request, а также традиционных методов - циклической балансировки, наименьшей загрузки и др.;
- устойчивая балансировка на 7 уровне модели OSI, контролируется программным способом, используя технологии OverDrive.

Для автоматизации управления вычислительными ресурсами системы используется программное обеспечение Hewlett-Packard OpenView Operations и Performance (OVOP) и Microsoft SMS. Вся собранная информация о состоянии вычислительной платформы фильтруется, регистрируется, преобразуется в единый формат и отправляется в управляющий сервер и фиксируется в исторической базе данных. В итоге средствами OVOP отслеживается доступность и производительность основных ресурсов вычислительных платформ: процессоров, памяти, дисковых подсистем, сетевых интерфейсов, файловых систем, процессов и сервисов ОС, а также пользователей. Средства SMS активно используются для контроля состояния и управления программным обеспечением и настройками рабочих станций сотрудников ЦОД.

Для надежного функционирования оборудования, обеспечены соответствующие условия окружающей среды. Для обеспечения «климатического режима» использованы кондиционеры «ACCURATE» с непосредственным испарением хладагента и воздушным охлаждением типа AX, использован фальшпол - систему съемных плит и решеток на регулируемых пьедесталах для обеспечения доступа к пространству под плитами.

Система раннего обнаружения пожара и пожаротушения реализована на базе построена на базе системы Vesda. В состав комплексной системы безопасности входят системы охранной сигнализации, видеонаблюдения, контроля и управления доступом, специальные средства. Система безопасности обеспечивает централизованное управление всеми подсистемами как с единого, так и со специализированных пультов, мониторинг состояния систем в режиме реального времени; ведение протоколов событий систем с генерацией отчетов по запросу; автоматическое оповещение сотрудников о внештатных ситуациях.