**Отчёт – 2011**

**Основные направления деятельности**

Основное направление работ лаборатории связано с исследованиями в области применения вычислительных кластеров для решения больших задач, а также разработкой и эксплуатацией современных высокопроизводительных вычислительных средств, предоставлением вычислительных ресурсов и консультационного обслуживания пользователям ССКЦ КП:

* Исследования и разработка методов управления потоком задач ЦКП на основе методов имитационного моделирования (совместно с лаб. моделирования процессов в сетях).
* Создание методов и инструментария для анализа и отладки алгоритмов и программ управления исполнением больших параллельных программ в кластерах экзафлопсного уровня с использованием методов имитационного моделирования (совместно с лаб. моделирования процессов в сетях, лаб. методов Монте-Карло).
* Экспериментальная проверка разработанных методов, допускающих высокую степень масштабируемости на кластерах ССКЦ СО РАН и МГУ (совместно с лаб. моделирования процессов в сетях, лаб. методов Монте-Карло).
* Обеспечение вычислительными ресурсами работ институтов СО РАН и университетов Сибири по математическому моделированию в фундаментальных и прикладных исследованиях, включая биоинформатику.
* Координация работ по развитию суперкомпьютерных центров Сибири, осуществляемая Советом по супервычислениям при Президиуме СО РАН.
* Организация обучения специалистов СО РАН и студентов университетов (ММФ и ФИТ НГУ, НГТУ) методам параллельных вычислений на суперкомпьютерах (поддержка ежегодных зимних и летних школ по параллельному программированию для студентов).
* Сотрудничество с INTEL, HP и промышленными организациями, тестирование новых процессоров.
* Сетевое взаимодействие с другими Суперкомпьютерными центрами СО РАН, Москвы и других городов России, а также зарубежных стран, совместная разработка технологий распределенных вычислений.

**Важнейшие достижения**

В соответствии с планом работ в 2011 г. развивались два основных направления: наращивание вычислительных мощностей ССКЦ и проект объединения вычислительных ресурсов СО РАН: ССКЦ, НГУ, ИВТ, ИЦиГ, ИХБиФМ, ИЯФ, по схеме «звезда» с использованием 10 G-ного канала (данная работа выполнялась совместно с ИВТ, НГУ и ИЯФ СО РАН).

**Вычислительные ресурсы ССКЦ КП СО РАН**

1. Кластерный суперкомпьютер **НКС-30Т** в составе:
* два шасси *HP* BladeSystem *c7000 Enclosure*, 32 сервера сверхплотной упаковки (двойные «лезвия») hp ProLiant BL2x220c **G5** (64 вычислительных модулей по два 4-х ядерных процессора Intel Quad-Core Xeon Е5450 и 16 ГБайт  оперативной памяти),
* четыре шасси *HP* BladeSystem *c7000 Enclosure*, 64 сервера сверхплотной упаковки (двойные «лезвия») hp ProLiant BL2x220c **G6** (128 вычислительных модулей по два 4-х ядерных процессора Intel Quad-Core Xeon Е5540 и 16 ГБайт  оперативной памяти),
* три шасси *HP* BladeSystem *c7000 Enclosure*, 48 серверов сверхплотной упаковки (двойные «лезвия») hp ProLiant BL2x220c **G7** (96 вычислительных модулей по два 6-ти ядерных процессора Intel Quad-Core Xeon Е5670 и 24 ГБайта  оперативной памяти),
* расширение кластера серверами с **GPU Nvidia**: 40 серверов SL390s G7 с тремя GPU M2090 на каждом. Пиковая производительность на GPU 79,800 ТФлопс, а с учётом серверов SL390s - 85,425 ТФлопс,
* параллельная файловая система **IBRIX** полезной емкостью 32 Тбайта дисковой памяти, которая поддерживается 4-мя серверами HP DL380 G6 и работает по InfiniBand.

Пиковая производительность **НКС-30Т** составляет **30** **ТФлопс**, производительность на тесте HPL (A Portable Implementation of the High-Performance Linpack Benchmark for Distributed-Memory Computers)  – 21,92 ТFlops. В декабре после ввода в эксплуатацию гибридного расширения кластера с GPU Nvidia суммарная пиковая производительность НКС-30Т составит **115.8 ТФлопс**.

1. Кластерный суперкомпьютер  **НКС-160** в составе 84 вычислительных модуля hp Integrity rx1620 (два процессора Intel Itanium2, 4 ГБайт оперативной памяти). Пиковая производительность НКС-160 составляет **1075 ГФлопс**, производительность на тесте HPL (A Portable Implementation of the High-Performance Linpack Benchmark for Distributed-Memory Computers)  – 828.7 GFlops
2. Сервер с общей памятью HP ProLiant DL980 **G7** с 4-мя 10-ти ядерными процессорами Intel Е7-4870 с тактовой частотой 2.4 ГГц, оперативной памятью **512 ГБайт** и 8 SAS дисками по 300 ГБайт. Пиковая производительность сервера в текущей конфигурации составляет 384 ГФлопс.
3. Сервер с общей памятью hp ProLiant DL580 **G5** в составе 4-х процессоров Intel Quad-Core Xeon X7350 (4 ядра) и  **256 ГБайт** оперативной памяти с дисковым массивом HP Storageworks SFS20, имеющим 9 ТБайт “сырой” дисковой памяти.
4. Сервер с общей памятью hp Integrity rx4640-8 в составе 4-х процессоров Intel Itanium2 и **64 ГБайт** оперативной памяти.

**Инфраструктура машинного зала ССКЦ**

Увеличение вычислительных ресурсов при реализации разрабатываемых проектов увеличит энергопотребление компьютерного оборудования  более чем 500 КВт. Это потребует такой же мощности источник бесперебойного электропитания и систему охлаждения с такой же производительностью по холоду. Обе эти системы должны иметь достаточный резерв, так как выход из строя любой из них приведёт к остановке всего компьютерного оборудования.

Ориентировочная стоимость системы водяного охлаждения с применением локальных кондиционеров с резервированием по схеме N+1 и без резервирования чиллера на 260 КВт составляет 10 млн. руб.  Создание таких достаточно больших систем обеспечения надёжной работы вычислительного оборудования требует привлечение опытных и квалифицированных проектировщиков. Ориентировочная стоимость проекта системы охлаждения на 500 КВт по холоду составляет более 1 млн. руб.

В 2011 году закуплен и введен в эксплуатацию промышленный прецизионный кондиционер BlueBox Datatech UEDA с производительностью по холоду 60 KВт.

В конце 2011 года закуплен и в настоящее время находится в стадии запуска промышленный прецизионный кондиционер BlueBox с гидромодулем и производительностью по холоду 75 KВт; основная его особенность это поддержка режима FreeCooling.

Мощность источника бесперебойного питания (ИБП) APC Symmetra PX2 увеличена до 160 КВт за счёт покупки дополнительных модулей.

Общее энергопотребление центра обработки данных (ЦОД) ССКЦ составляет **294** КВт.

**Программное обеспечение/инструментальные средства разработки**

Кластера ориентированы для решения параллельных задач с использованием Message Passing Interface (MPI), а многопроцессорные сервера с общей памятью Symmetric MultiProcessor (SMP) на решение задач, прежде всего 3-D, требующих большой оперативной памяти, а также параллельных задач с использованием OpenMP. При таком подходе поддерживаются две современных парадигмы параллельных вычислений – MPI для систем с распределенной памятью (кластеров) и OpenMP для систем с общей памятью. Гибридная схема предусматривает запуск на каждый вычислительный узел кластера по одному MPI-процессу, который запускает внутри каждого вычислительного модуля  несколько потоков с помощью OpenMP.

На кластере **НКС-30Т** установлен Intel MPI 4 и Intel TraceAnalizer / Collector, компиляторы Intel C++ и Intel Fortran Composer XE for Linux Version 2011 Update 5, включающие в себя библиотеки Intel MKL, Intel IPP и Intel TBB.

На НКС-30Т также установлены параллельные версии Gromacs, Quantum Espresso и Bioscope.

Для программирования на **GPU Nvidia** существуют наборы инструментов для CUDA C/C++/Fortran, OpenCL, DirectCompute, NVIDIA Parallel Nsight™ для Visual Studio. Доступны также Matrix Algebra on GPU and Multicore Architectures (MAGMA, <http://icl.cs.utk.edu/magma/index.html>) GPU-accelerated linear algebra library that utilizes the NVIDIA CUDA parallel computing architecture (CULA, <http://www.culatools.com/>).

На кластере **НКС-160** установлена версия MPI mvapich 1.2.6, компиляторы
Intel C++ и Intel Fortran for Linux версии 10.1 и библиотека Intel MKL 10.0.

На НКС-160 также установлены коммерческие пакеты Fluent 6.3 и Gaussian 03 и свободно распространяемые (с открытым исходным кодом) LAMMPS Molecular Dynamics Simulator, Gromacs и Quantum Espresso. Хотя архитектура Intel Itanium2 уже не применяется в области высокопроизводительных вычислений, но установленные пакеты Fluent 6.3 и Gaussian 03 делают его привлекательным для пользователей.

На многопроцессорном сервере с общей памятью ProLiant DL580 **G5** также установлены  компиляторы Intel C++ и Intel Fortran Composer XE for Linux Version 2011 Update 5, включающие в себя библиотеки Intel MKL, Intel IPP и Intel TBB. Компиляторы Intel поддерживают многопоточность и OpenMP. Одинаковый комплект базового программного обеспечения на кластерах и серверах упрощает работу пользователей.

На сервере с общей памятью hp Integrity rx4640-8 установлены компиляторы Intel C++ и Intel Fortran for Linux версии 11.1 и библиотека Intel MKL 10.2.

Кроме ежегодного приобретения лицензий на системное ПО и средства разработки (Intel SW, PBS Pro, RedHat Linux и т. п.) для обеспечения задач моделирования на НКС-30Т необходимо приобрести прикладные коммерческие пакеты программ: **STAR CD, Ansys Multiphisycs, Ansys LS-DYNA, Fluent, Gaussian 09**. Ориентировочная стоимость перечисленных пакетов составляет 15 млн. руб.

**Некоторая статистика**

Использование процессорного времени ССКЦ в 2011 г. (час.)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Организации** | **НКС-160** | **НКС-30T** | **∑** | **%** |
| **ИЦиГ** |   | 2401720,56 | **2401720,56** | 47,65 |
| **ИВМиМГ** | 8006,70 | 888407,04 | **896413,74** | 17,79 |
| **ИХиХТ (Красноярск)** | 126673,66 | 380176,56 | **506850,22** | 10,06 |
| **ИК** | 42701,40 | 400436,88 | **443138,28** | 8,79 |
| **НГУ** | 12157,30 | 134954,88 | **147112,18** | 2,92 |
| **ИХКиГ** | 37188,62 | 88826,16 | **126014,78** | 2,50 |
| **ИТПМ** | 59904,45 | 61830,24 | **121734,69** | 2,42 |
| **ИЯФ** | 25591,94 | 59264,64 | **84856,58** | 1,68 |
| **ИТ** | 176,74 | 62445,84 | **62622,58** | 1,24 |
| **СИСТЕМА** | 0,19 | 56874,24 | **56874,43** | 1,13 |
| **ИАТЭ (Обнинск)** |   | 54518,88 | **54518,88** | 1,08 |
| **ИХБиФМ** |   | 33858,96 | **33858,96** | 0,67 |
| **ИНХ** |   | 20860,80 | **20860,80** | 0,41 |
| **НГТУ** | 18734,13 | 89,76 | **18823,89** | 0,37 |
| **ИВТ** |   | 17819,28 | **17819,28** | 0,35 |
| **ИНГиГ** | 1316,84 | 10297,92 | **11614,76** | 0,23 |
| **НИЦЭВТ (Москва)** |   | 10430,88 | **10430,88** | 0,21 |
| **ИКЗ (Тюмень)** | 1605,49 | 8327,04 | **9932,53** | 0,20 |
| **СибНИА** | 6730,63 |   | **6730,63** | 0,13 |
| **ИФП** |   | 4945,44 | **4945,44** | 0,10 |
| **ИМ** | 1572,42 | 4,32 | **1576,74** | 0,03 |
| **Компания Котес** | 916,51 |   | **916,51** | 0,02 |
| **ОФ ИМ (Омск)** |   | 437,28 | **437,28** | 0,01 |
| **«Вектор»** | 134,76 | 0,96 | **135,72** | 0,00 |
| **ЛИН (Иркутск)** |   | 1,44 | **1,44** | 0,00 |
| **ИТОГО** | **343411,76** | **4696530** | **5039941,76** | 100,00 |

Проекты и программы, при выполнении которых использовались услуги ССКЦ в 2011 г.

|  |  |
| --- | --- |
| Всего грантов, программи проектов — **131**Из них Российских — **131**,Международных — 0.Грантов РФФИ – **44**Программ РАН – **26**Проектов СО РАН – **31**Программ Минобразнауки – **4**Другие – **26**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Всего публикаций – **173**Российских – **98**Зарубежных –**75** | **Проекты по институтам:**ИВМиМГ – **34**ИЦиГ – **18**ИТПМ – **15**ИТ – **12**ИК – **11**ИХКиГ – **9**ИНГиГ – **8**ИКЗ (Тюмень) – **3**ИХиХТ (Красноярск) – **3**СибНИА – **3**ИВТ – **2**ИМ – **2**ИНХ – **2**«Вектор» – **2**ИХБФМ – **2**ИЯФ – **2**ИФП – **1**ОАО "НИЦЭВТ" (Москва) – **1**ОФ ИМ (Омск) – **1** |

Области исследований задач, решаемых в ССКЦ (2011 г.):

1. Информационно-телекоммуникационные технологии –
ИВМиМГ, ИВТ, НИЦЭВТ (Москва)
2. Вычислительная математика – ИВМиМГ, НГУ
3. Математическая физика – ИВМиМГ, ИМ, НГУ
4. Химическая физика – ИХКиГ
5. Геофизика – ИВМиМГ, ИНГиГ, ИТ
6. Астрофизика - ИК
7. Физика высоких энергий – ИЯФ
8. Биотехнологии – ИЦиГ, ИХБиФМ, ИТ, «ВЕКТОР» , ОФ ИМ (Омск)
9. Нанотехнологии – ИВМиМГ, ИФП, ИНХ, ИХиХТ (Красноярск)
10. Квантовая химия – ИК
11. Астрохимия - ИК
12. Астрокатализ – ИК
13. Новые материалы и химические технологии – ИНХ, ИК
14. Материаловедение – ИКЗ (Тюмень)
15. Энергосберегающие технологии – ИКЗ (Тюмень)
16. Экология и рациональное природопользование – ИВМиМГ, ИНГиГ
17. Космические и авиационные технологии – ИТПМ, ИКЗ (Тюмень)
18. Аэродинамика – ИТПМ, НГТУ, СибНИА
19. Теплоэнергетика – ИТ

**План работ на 2012 год**

В настоящее время, при поддержке Приборной комиссии СО РАН, суммарная мощность ЦКП ССКЦ при ИВМиМГ СО РАН превысила 100 TFlops, однако эффективное использование данных мощностей зависит от программного обеспечения и пакетов прикладных программ, установленных на вычислительных кластерах. Отметим, что увеличение производительности было достигнуто в основном за счет закупки современного гибридного кластера, использование которого в ЦКП требует закупки и установки соответствующих версий пакетов: ANSYS 13 (CFD+CFX), PGI Accelerator Compilers (С/C++/Fortran + CUDA) и Gaussian 9.

В последнее время значительно возрос спрос на использование прикладного программного пакета **ANSYS Fluent,** установленного на кластере Сибирского Суперкомпьютерного Центра ИВМиМГ СО РАН. На данный момент наибольшим спросом пакет пользуется у сотрудников следующих институтов: Институт Катализа им. Г.К. Борескова СО РАН, Институт неорганической химии им.А.В. Николаева СО РАН, Институт теплофизики им. С.С. Кутателадзе СО РАН, Институт теоретической и прикладной механики им. С.А. Христиановича СО РАН, Институт Вычислительной Математики и Математической Геофизики СО РАН.

На настоящий момент на кластере ССКЦ НКС-160 установлен пакет Gaussian 3.

1. Существующих пользователей уже не вполне удовлетворяет производительность старого кластера, поэтому этот пакет следует поставить и на новый кластер, но перенос невозможен (на новый кластер возможна только покупка новой версии).

2. Существующих пользователей также не удовлетворяет и функционал старой версии пакета.

3. Некоторые потенциальные пользователи не становятся реальными из-за старой версии и низкой производительности. После покупки новой версии и установки ее на новый кластер количество пользователей (а, возможно, институтов) увеличится.

Последняя версия **Gaussian 9** эффективно использует графические ускорители для вычислений, что позволит значительно поднять производительность, благодаря 120 графическим ускорителям последнего апгрейда вычислительного кластера НКС-30Т ССКЦ.

В связи с закупкой 40 серверов HP SL390s G7 с графическими ускорителями (GPU) NVidia Tesla M2090 (пиковая производительность закупленных серверов порядка 80ТФлопс) необходимы специализированные компиляторы, использующие возможности графических ускорителей. В частности пакета **PGI Accelerator Compilers (C/C++/Fortran + CUDA).** На данный момент хорошо отработана схема программирования под графические ускорители на языке C++. Однако, большинство вычислительных разработок в СО РАН написаны на языке Fortran. Самый лучший компилятор с языка Fortran, с возможностью использования GPU, поставляется компанией Portland Group. Версия компиляторов PGI Accelerator также позволяет пользователям писать на высокоуровневом языке, не используя достаточно сложные конструкции копирования данных в память ускорителя, синхронизации процессов и других, что значительно сокращает время освоения программирования с использованием графических ускорителей.

Поэтому основное внимание в 2012 году будет уделено закупке, установке и освоению пакетов и программного обеспечения для кластеров ССКЦ.

Для того чтобы скорость работы с параллельной файловой системой IBRIX не лимитировала работу расширенного кластера, в  том числе на задачах Биоинформатики - необходимо увеличить число серверов IBRIX минимум вдвое. При этом скорость обменов будет расти пропорционально числу серверов IBRIX, т.е. тоже возрастёт вдвое. Для эффективного использования вычислительных мощностей, считаем необходимым увеличить производительность кластерной СХД как минимум вдвое. Ориентировочно наращивание параллельной файловой системы IBRIX будет стоить порядка 6.0 - 6.5 млн. руб.

Планируется нарастить сервер с общей памятью HP ProLiant DL980 G7 до максимальной конфигурации для чего необходимо закупить еще 4 10-ти ядерных процессора Intel Е7-4870 с тактовой частотой 2.4 ГГц и оперативной памятью
**1536 ГБайт**.

**ГРИД – сегмент СО РАН**

В 2011 году ССКЦ СО РАН включён в договор о Суперкомпьютерной сети ННЦ СО РАН, первоначально заключенного между НГУ, ИВТ СО РАН и ИЯФ СО РАН в 2010 г. Работы по Грид - сегменту СО РАН ведутся по следующим направлениям:

1. проект по созданию ГРИД-сайта ННЦ СО РАН в рамках Национальной нанотехнологической сети (ГридННС);
2. работа в рамках ЦКП «Биоинформатика», в состав которого входит ИВМиМГ СО РАН;
3. выполнения работ в области высокопроизводительных вычислений в физике высоких энергий и обработки данных физических экспериментов, осуществляемых в ИЯФ СО РАН (эксперименты КЕДР, СНД, КМД-3).

В рамках подключения к ГридННС запланирована установка программного обеспечения на выделенный сервер (ГРИД - ШЛЮЗ), требования к которому описаны на странице <http://www.ngrid.ru/docs/trunk/installation_administration/i_grid_gateway.html>

Для обеспечения доступа сервера к параллельной файловой системе кластера IBRIX закуплен дополнительный InfiniBand адаптер.

Для выполнения работ в области высокопроизводительных вычислений в физике высоких энергий и обработки данных физических экспериментов, осуществляемых в ИЯФ СО РАН (эксперименты КЕДР, СНД, КМД-3) в частности запланировано использование ГРИД – ШЛЮЗ, который будет подключен по сети 10 Гбит/с к суперкомпьютерной сети ННЦ СО РАН, установка виртуальной машины на выделенные блейд-сервера кластера НКС-30Т и другого необходимого программного обеспечения. В настоящее время для выхода по 10 Гбит/с к суперкомпьютерной сети ННЦ временно используется управляющий узел НКС-30Т, что в принципе ограничивает его использование в качестве полноценного ГРИД – ШЛЮЗа.