Независимая
научно-практическая конференция
«Разработка ПО 2011»
31 октября - 3 ноября, Москва

Моделирование компьютерного кластера на распределённом симуляторе

Речистов Григорий, Павел Шишпор, Александр Иванов

Московский физико-технический
институт

Введение

- Описываются характеристики кластера МФТИ, развиваемого для задач вычислительной биологии
- Симуляционная модель, построенная для анализа «узких мест» в его производительности
- Результаты измерений на модели, их валидация.

Проблемы симуляционного подхода

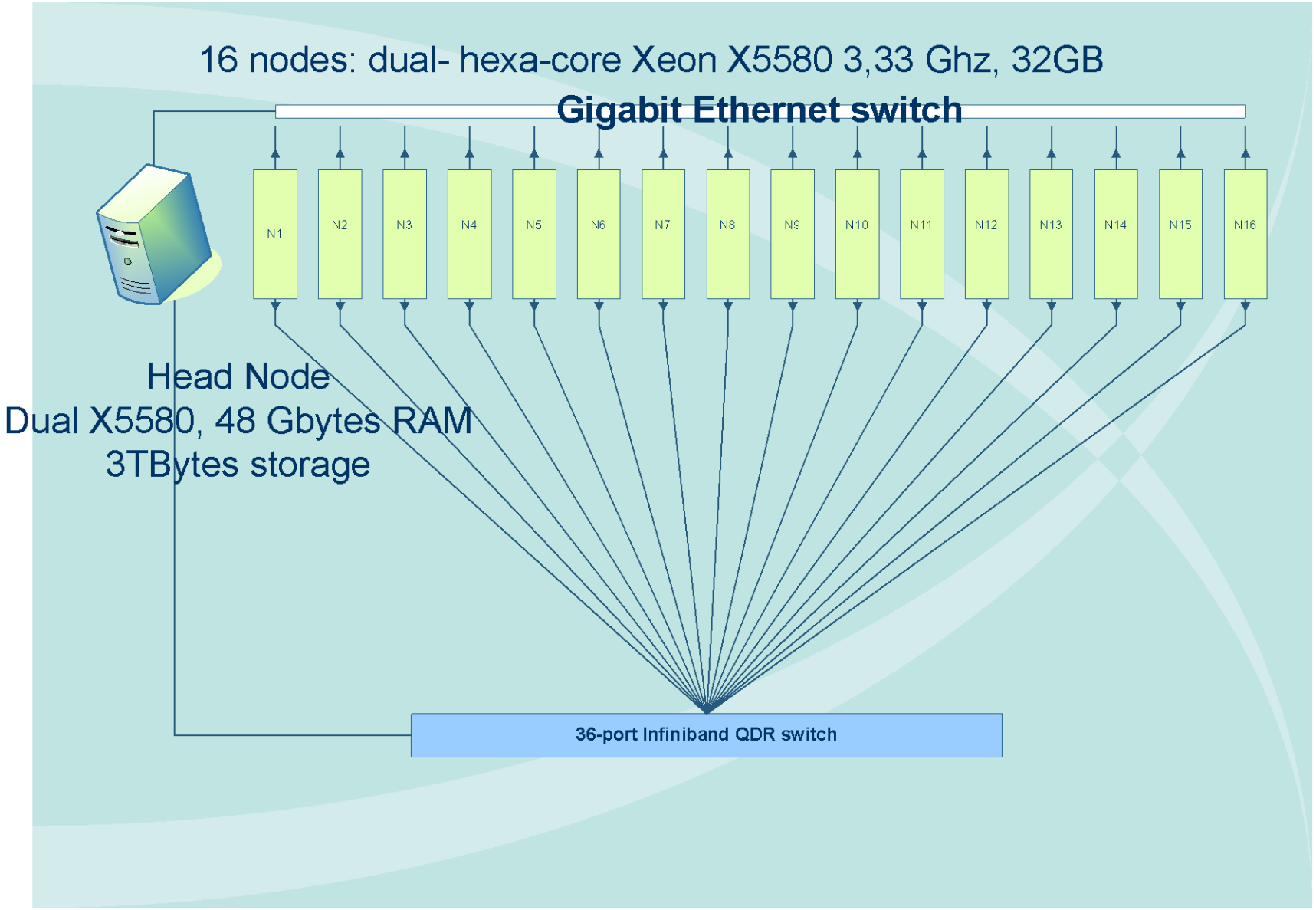
- Малая скорость при последовательном моделировании => необходимы параллельные модели
- Ресурсов одной машины также не хватит для воплощения модели целого кластера, поэтому она должна быть распределённой, т.е. сама выполняться на (меньшем) кластере.

Характеристики кластера

Число вычислительных узлов	16
Процессор головного узла:	Intel Xeon 5580 3,33 ГГц (2шт х 6 ядер)
Объём памяти головного узла	48 Гбайт
Дисковое хранилище головного узла	3 Тбайт
Процессоры выч. узлов	Intel Xeon 5580 3,33 ГГц (2шт х 6 ядер)
Объём памяти каждого вычисл. узла	32 Гбайт

Полное число вычислительных ядер 192

Полный объём памяти для вычислений 512 Гбайт


16 nodes: dual- hexa-core Xeon X5580 3,33 Ghz, 32GB
Gigabit Ethernet switch
N1
N2
N3
N4
N5
N6
N7
N8
N9
N10
N11
N12
N13
N14
N15
N16
Head Node
Dual X5580, 48 Gbytes RAM
3TBytes storage
36-port Infiniband QDR switch

Simics

- Полносистемная симуляция
- Множество моделей устройств, позволяющих быстро создавать прототипы систем.
- Лёгкость разработки и подключения новых моделей устройств.
- Средства автоматизации процесса симуляции и измерений с помощью сценариев.
- Использование двоичной трансляции и аппаратных расширений виртуализации Intel VTx.
- Многопоточное исполнение.
- Распределённый режим работы.

На чём запускалась модель

- Intel Xeon(R) 5150 (Woodcrest), 8 ядер x 2 HT @ 2.66 ГГц
- ОЗУ – 40 Гбайт
- 2 Гб HDD для образов дисков.

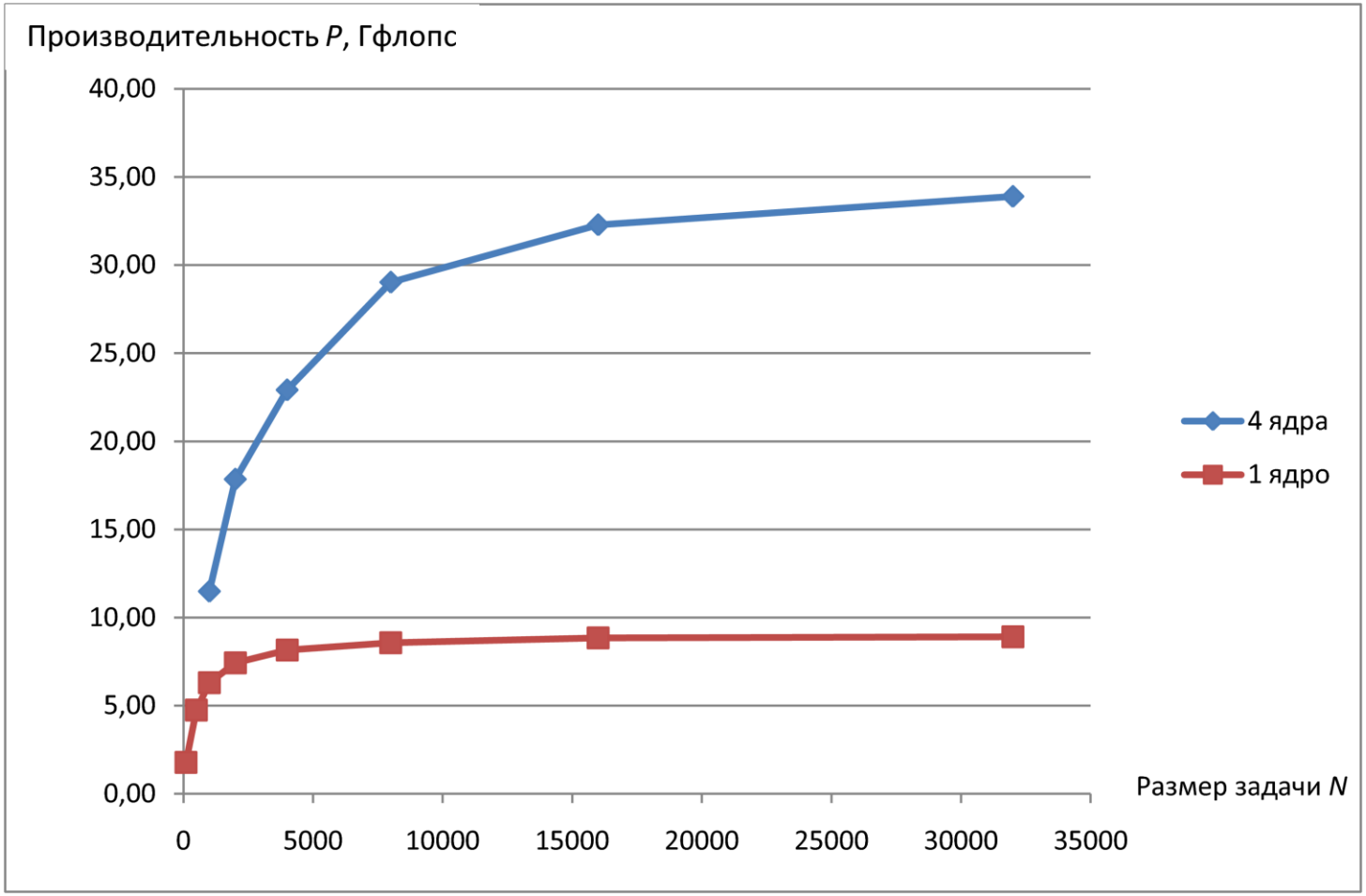
- Модель работает с высокой скоростью:
  - Linux на всех узлах поднимается за 8 минут
  - Linpack с размером матрицы *N* = 2000 проходит за 30 минут.

Задачи

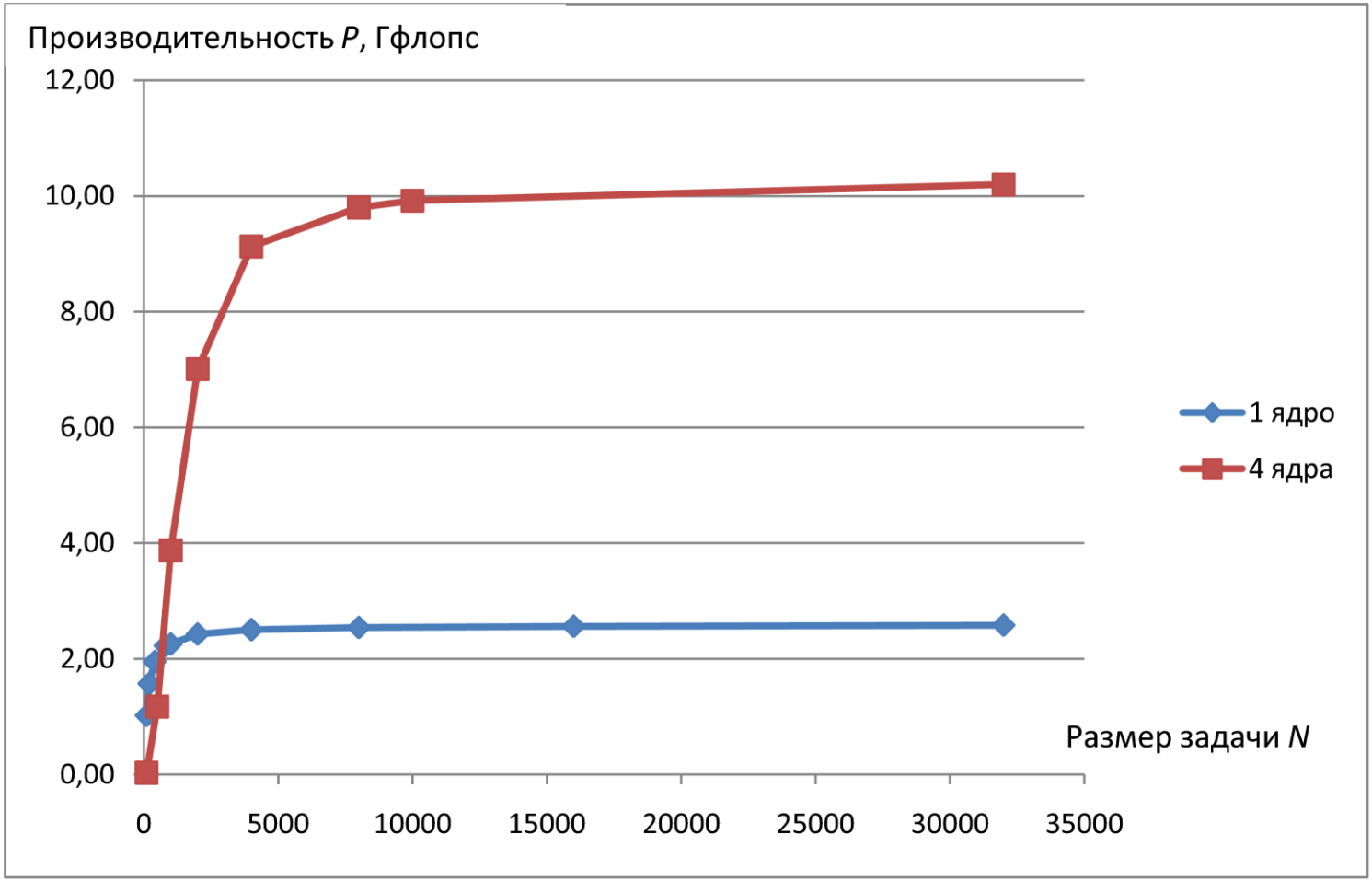
- High Performance Linpack

- Изучались
  - сообщаемая бенчмарком производительность (FLOPS)
  - характер сетевого взаимодействия MPI-процессов (с помощью интеграции Simics - Wireshark)

Результаты Linpack (реальная аппаратура)



Результаты Linpack (Simics)



Нагрузка на сетевой коммутатор на модели кластера.

Linpack: N, PxQ	Нагрузка (Mbit/s)	Средний размер пакета (Bytes)	Gflops
4000, 1x192	**1300**	5800	0,56
4000, 6x32	**400**	3700	1,1
4000, 16x12	**150**	1100	0,46

Влияние задержки сетевого пакета на производительность Linpack в модели кластера.

Linpack N, PxQ	Gflops, Delay= 80мкс	Gflops, Delay= 160мкс	Gflops, Delay= 240мкс	Gflops, Delay= 320мкс	Gflops, Delay= 400мкс
4000, 1x192	0,56	0,55	0,56	0,57	0,56
4000, 6x32	1,1	1,1	1,1	1,1	1,1
4000, 16x12	4,6	4,6	4,6	4,5	4,5

Заключение

- Произведено исследование кластера в его текущей конфигурации
- Дальнейшие изменение модели для отражения развития системы (в 2012 г.)
- Проведение исследований приложений молекулярной динамики по отработанной методике.

Спасибо за внимание!