

Исследование масштабируемости задач вычислительной гидроаэродинамики на различных многоядерных и многопроцессорных архитектурах¹

В.А. Васильев, А.Ю. Ницкий

Проведено численное исследование влияния архитектуры кластера на эффективность решения задач вычислительной гидроаэродинамики (ВГАД) на примере задач течения турбулентной жидкости в тонком слое уплотнительных узлов гидромашин с использованием пакета ANSYS CFX. Получена существенная зависимость производительности от архитектуры и способа загрузки вычислительных узлов. Проведена оптимизация решения задач ВГАД на кластере “СКИФ Урал” ЮУрГУ.

Задачи вычислительной гидроаэродинамики требуют значительных вычислительных ресурсов и мощных программных средств, способных использовать возможности, предоставляемые современными вычислительными системами. При решении больших и сверхбольших задач важным фактором является оптимизация совместной работы вычислительной системы (кластера) и программного продукта. Цель данной работы – исследование эффективности суперкомпьютера СКИФ Урал ЮУрГУ, на базе четырехъядерных процессоров E5472 и определение необходимых и достаточных условий оптимизации решения совместных задач гидродинамики и динамики ротора на вычислительном кластере ЮУрГУ.

Для проведения вычислительных экспериментов в качестве тестовой задачи использованы задачи расчета гидродинамики тонкого турбулентного слоя жидкости в щелевом уплотнении мощного питательного насоса. Тесты ориентированы на использование известного параллельного пакета, широко используемого на различных суперкомпьютерных платформах – пакета инженерного анализа ANSYS[®] CFX v. 11.0. В отличие от использования тестов, таких как LINPACK, основанных на решении системы линейных уравнений прямыми методами, гидродинамическая задача большой размерности, как правило, решается итерационными методами и представляет собой сложную математическую модель реальных объектов и конструкций.

Расчеты проводились на кластере СКИФ Урал ЮУрГУ, имеющем LINPACK-производительность 12,2 TFlops. Технические характеристики приведены в таблице 1.

Таблица 1. Технические характеристики вычислительного кластера СКИФ Урал ЮУрГУ

Количество узлов / процессоров / zlth	166 / 332 / 1328
Процессоры вычислительного узла	2 x Intel Xeon E5472 (4 ядра по 3.0 GHz)
ОЗУ/ дисковая память вычислительного узла	8GB DDR3 / 120 GB
Общая дисковая память / система хранения	49,39 TB / Panasas ActiveStorage 5100
Тип системной сети	InfiniBand (20Gbit/s, макс. задержка 2 мкс)
Тип управляющей (вспомогательной) сети	Gigabit Ethernet
Сервисная сеть	СКИФ ServNet
Пиковая производительность/Linpack	16 TFlops / 12,2 TFlops
Операционная система	SUSE Linux Enterprise Server 10 , Windows CCS
Тестируемое программное обеспечение	ANSYS [®] CFX Academic Research, v. 11.0, SP1
Библиотека MPI (ОС Linux)	HP MPI v.02.02.05.01
Система бесперебойного электропитания	APC Symmetra 160 kVA

¹ Авторы статьи выражают благодарность официальному представителю ANSYS компании "Делкам-Урал" за предоставленные на время проведения тестирования HPC лицензии CFX v11.0 SP1

Также для расчетов использовалась рабочая станция Supermicro, представляющая собой 2-х процессорную ЭВМ с объемом ОЗУ 32 Гбайт на базе процессоров AMD Opteron. Технические характеристики ЭВМ приведены в таблице 2. Кроме того, ряд тестов на малой задаче был проведен на рабочей станции Supermicro с двумя четырехъядерными процессорами Intel Xeon E5520@2,27 ГГц и 24 Гбайт DDR3 ОЗУ.

Таблица 2. Технические характеристики рабочей станции

Количество процессоров	2
Количество вычислительных ядер	4
Тип процессора	2 x AMD Opteron 2216 (Santa Rosa, 2 ядра по 2.4 GHz) 2 x AMD Opteron 2354 (Barcelona, 4 ядра по 2.2 GHz) 2 x AMD Opteron 2427 (Istanbul, 6 ядер по 2.2 GHz)
ОЗУ	32GB DDR2
Дисковая память вычислительного узла	144 GB
Операционная система	Windows XP Pro x64 SP2, SUSE LES 10
Тестируемое программное обеспечение	ANSYS® CFX Academic Research, v. 11.0, SP1
Библиотека MPI	MPICH2

С целью оценки влияния размера задачи на эффективность использования вычислительной мощности кластера анализ проводился на трех задачах разного размера. **Малая задача** – 46 218 узлов (29 204 элементов), **средняя задача** – 328 392 узлов (242 663 элементов) и, наконец, **большая задача** – 9 903 873 узлов (9 871 040 элементов).

Малая задача представляет собой расчет гидродинамики тонкого турбулентного слоя в радиальной щели стендовой установки в двумерной осесимметричной постановке. Проводилось исследование течения жидкости в радиальном зазоре, образуемым диском, закрепленным на свободном конце модельного ротора стендовой установки (рис. 2). Диск диаметром 130 мм образует с корпусом радиальную щель зазором от 200 до 500 мкм. Внутренний радиус щели 66мм, внешний 93мм, длина радиального зазора 27 мм. На вход радиальной щели подавалась жидкость – вода с температурой порядка 40°С. На рис. 3. представлена двумерная осесимметричная модель радиальной щели в ANSYS® CFX-Pre и фрагмент расчетной сетки.

Средняя задача представляет собой расчет гидродинамики уравнивающего устройства питательного насоса СВПТ 1150, проводимый для определения упругих и демпфирующих свойств уплотнений проточной части, в частности, уплотнений разгрузочного устройства. Расчеты полей давления, полей скоростей в щелевом уплотнении уравнивающего устройства насоса СВПТ 350-850 проводились для цилиндрической щели диаметром 180 мм, длиной 185мм, радиальной щели, наружный диаметр которой составлял 300мм и внутренний 250 мм. Зазор цилиндрической и радиальной щелей принят соответственно 350 и 120 мкм. Перекачиваемая среда - питательная вода, температура 170°С, давление на входе в цилиндрическую щель 34,4 МПа, давление на выходе из радиальной щели 2,15 МПа.

Задача решалась в осесимметричной постановке (рис. 4) с заданием граничных условий по давлению на входе и выходе. Сетка строилась на одноградусном секторе, размер сетки составил 328392 узлов, 242663 элементов.

Большая задача представляет собой расчет гидродинамики тонкого турбулентного слоя в радиальной перекошенной щели стендовой установки. В отличие от первого варианта расчета (малая задача), в котором зазор в щели параллелен, решение данной задачи возможно лишь в трехмерной постановке. Это позволяет получить распределения давления по длине щели, приходящие конфузурной и диффузурной геометрии зазора (рис.5.).

Для перекошенной щели сетка построена из двух расчетных подобластей (Domain), которые затем стыкуются с помощью интерфейсов (Domain Interfaces), как показано на рис 6. Общее число узлов на перекошенной радиальной щели составило 9 903 873. Необходимая оперативная память 31,6 Гбайта, время счета на 16 ядрах кластера Скиф Урал составляет порядка 4 часов.

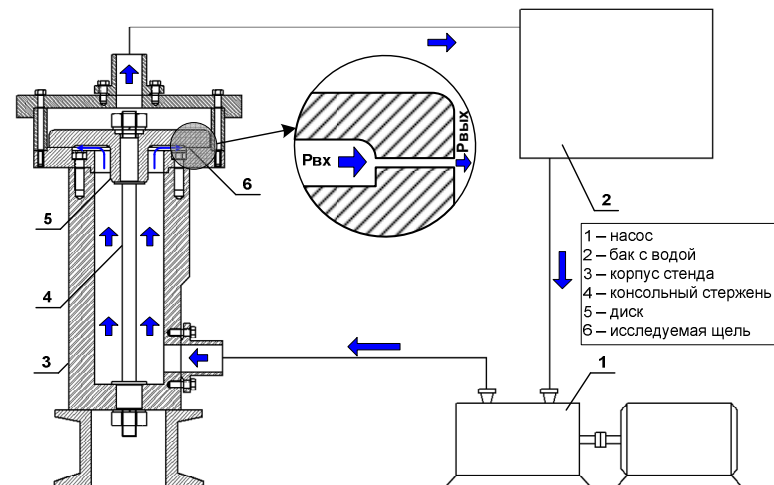


Рис. 2. Принципиальная схема стендовой установки и модельного ротора

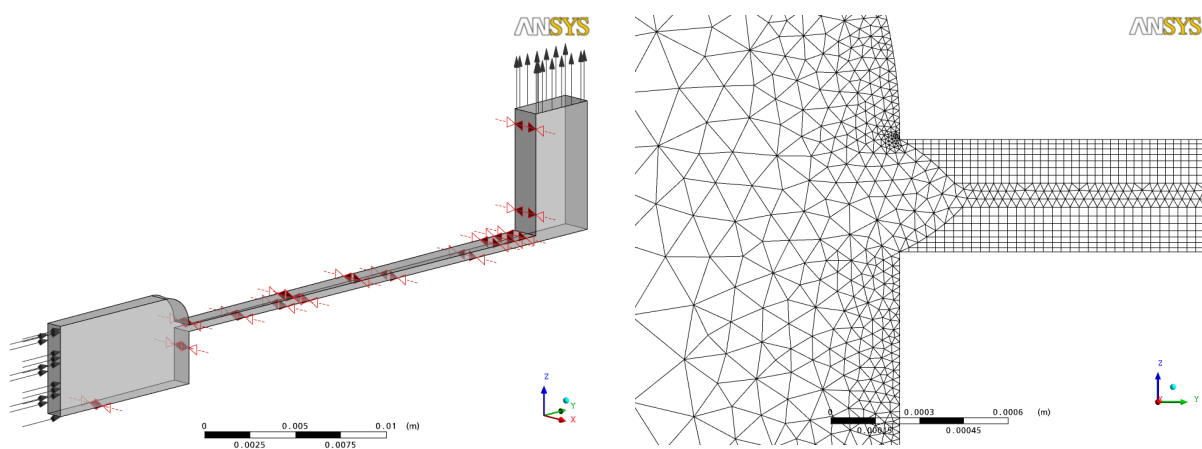


Рис.3. Модель малой задачи в ANSYS CFX-Pre и расчетная сетка

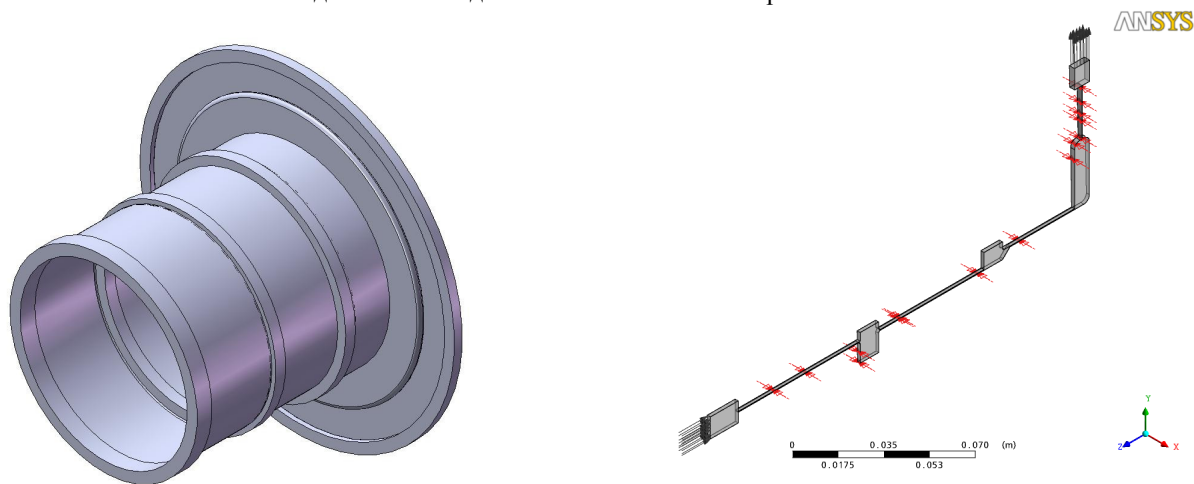


Рис.4. Геометрия щели уравновешивающего устройства питательного насоса СВПТ 350-850

В большинстве вариантов расчетов использовалась версия решателя CFX с двойной точностью чисел с плавающей точкой (double), кроме того, проведен ряд вычислительных экспериментов на решателе с одинарной точностью на тех задачах, где одинарная точность давала устойчивую сходимость решения.

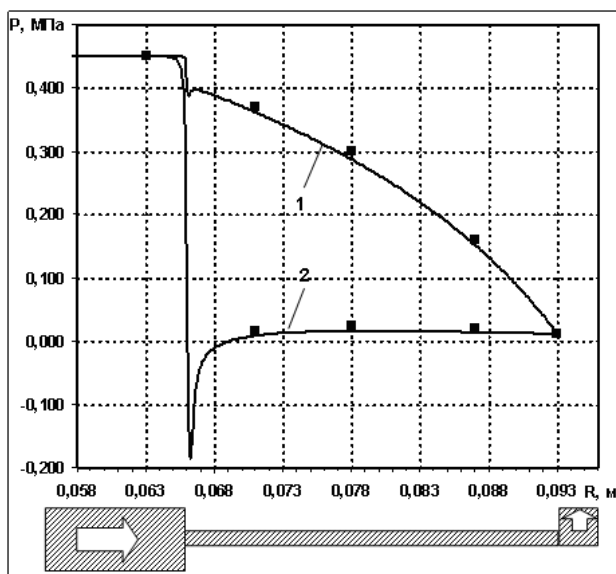


Рис. 5. Распределение давления по длине щели в сечениях конфузор (1), диффузор (2)

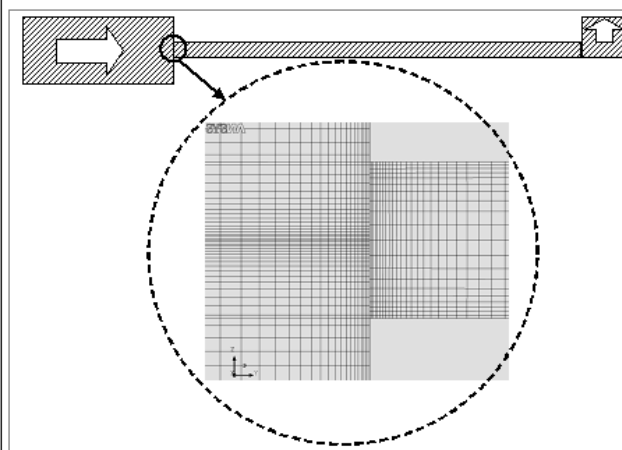


Рис. 6. Участок расчетной сетки на стыке 2х доменов: входной зоны и щелевого уплотнения

Результаты сравнительного анализа эффективности вычислений

Эффективность расчета, в значительной степени, зависит от процесса генерации сетки и оптимизации процесса распараллеливания сетки для последующего использования мощных кластерных систем. Эффективность разделения сетки при использовании пакета ANSYS CFX зависит от: метода разделения и числа блоков сетки при разделении.

В свою очередь, эффективность работы в параллельном режиме зависит также от размера задачи или числа узлов сетки, типа элемента, используемого при генерации сетки. Например, для тетраэдрических сеток рекомендуется не использовать меньше 30 000 узлов на блок сетки (на вычислительное ядро), для гексаэдрических сеток минимальное количество узлов на блок рекомендовано не менее 75 000 узлов [1]. Количество узлов на процессор не является жестким; при решении реальной задачи на определенной аппаратной платформе критическое количество узлов может быть больше или меньше.

Кроме того, узел кластера, содержащий два процессора и несколько ядер на каждом, может иметь ограничение по масштабированию из-за недостаточной пропускной способности шины памяти. По существу два центральных процессора могут потребовать больше ресурсов доступа к памяти, чем может обеспечить шина памяти.

Масштабируемость задачи на кластере в зависимости от числа используемых вычислительных ядер

Вычисления проводились на большой задаче с числом узлов равным 9 903 873. Результаты приведены на рис. 7 и в табл. 3. Эффективность вычислений оценивалась по числу итераций в час. Получены две кривые: первая с насыщением на уровне 200 ядер, вторая кривая показывает рост производительности до 300 ядер. Отличие в эффективности использования кластера зависит от количества используемых ядер вычислительных узлов. Первая кривая соответствует использованию полной вычислительной мощности каждого узла с загрузкой всех восьми ядер. Вторая кривая соответствует специальной загрузке, при которой на каждом отдельном узле загружается только 4 ядра. Более эффективным является второй вариант загрузки кластера. Это свидетельствует о вероятных проблемах, связанных с совместной работой процессоров на узле кластера. В ходе проведения вычислительных экспериментов было выявлено, что при использовании от 1го до 4х (включительно) ядер на узел производительность вычислений зависит от общего числа используемых ядер и не зависит от количества используемых ядер на вычислительном узле. На рис 7. и далее зеленой линией показано линейное ускорение.

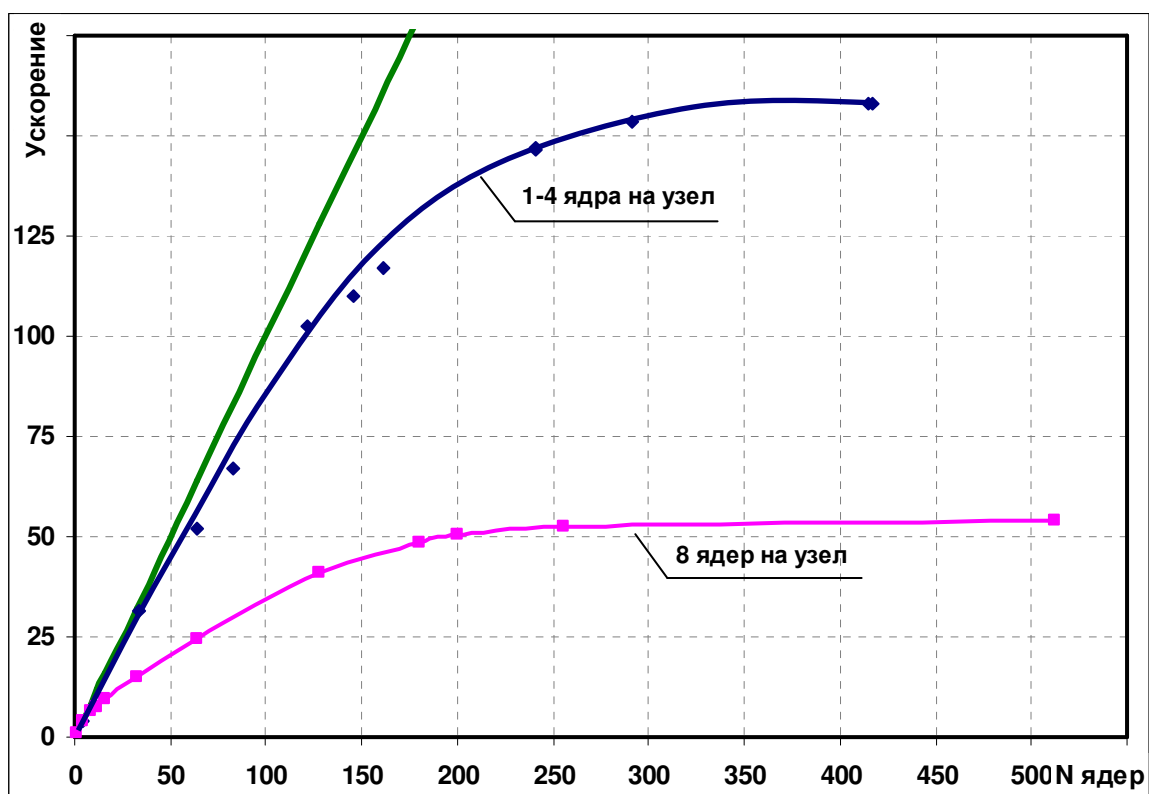


Рис. 7. Масштабирование большой задачи в зависимости от использования числа ядер внутри узла

Таблица 3. Масштабирование большой задачи

Кол-во ядер	Способ запуска	Время 100 итер.	Итер/час	Ускор.	Эфф-ть
417	104*4+1	0:13:08	457,0	157,8	37,8%
415	138*3+1	0:13:05	458,0	158,2	38,1%
291	145*2+1	0:13:31	444,4	153,5	52,7%
241	80*3+1	0:14:09	425,5	147,0	61,0%
241	120*2+1	0:14:08	424,6	146,6	60,8%
161	80*2+1	0:25:02	339,0	117,1	72,7%
146	146	0:18:47	319,1	110,2	75,5%
121	121	0:20:12	297,0	102,6	84,8%
83	83	0:31:03	193,5	66,8	80,5%
64	64	0:40:13	150,0	51,8	80,9%
33	33	1:05:33	91,6	31,6	95,9%
4	4	8:37:56	11,6	4,0	100,0%

Масштабируемость задач внутри вычислительного узла

Вычисления проводились на всех типах задач, производительность вычислений оценивалась по числу итераций в минуту. Результаты для малой задачи приведены на рис. 8. (узел кластера СКИФ Урал) и рис. 9. (рабочая станция AMD с процессорами Opteron 2354 и 2427, рабочая станция с двумя процессорами Intel E5520). При тестировании на узле СКИФ Урал получены две кривые: первая с насыщением на уровне 4х ядер при запуске задачи внутри узла, вторая кривая показывает больший рост производительности, и соответствует специальной загрузке, при которой к первому ядру одного узла последовательно подключаются по одному вычислительному ядру на каждом следующем узле, подключенному через сеть интерконнекта. Данный способ загрузки позволяет оценить "параллельный" потенциал задачи без ограничения архитектуры узла. Как оказалось, узел кластера, содержащий два процессора и восемь вычисли-

тельных ядер, имеет ограничение по масштабируемости, вероятнее всего, из-за недостаточной пропускной способности шины памяти. Подобное поведение наблюдается и на других пакетах ВГАД, использующих отличные от CFX алгоритмы решателей [2,3,4] и тестах производительности, основанных на решении систем линейных уравнений итерационными методами [5].

Аналогичная картина наблюдается и на рабочих станциях (рис. 9), однако общий уровень масштабируемости этих систем на малой задаче выше, чем предыдущей. Следует отметить, что новое поколение процессоров AMD Opteron Istanbul существенно не улучшило масштабируемость по сравнению с Barcelona, а линейка процессоров Intel Nehalem улучшила свои показатели по сравнению с предыдущей Harpertown и сравнялась по масштабируемости с системами AMD. Однако уровень абсолютной производительности систем оказался разным, и процессор E5520 значительно обгоняет конкурентов теста (рис.10), несмотря на значительное отставание по частоте от Xeon X5472 и по количеству ядер от Opteron 2427.

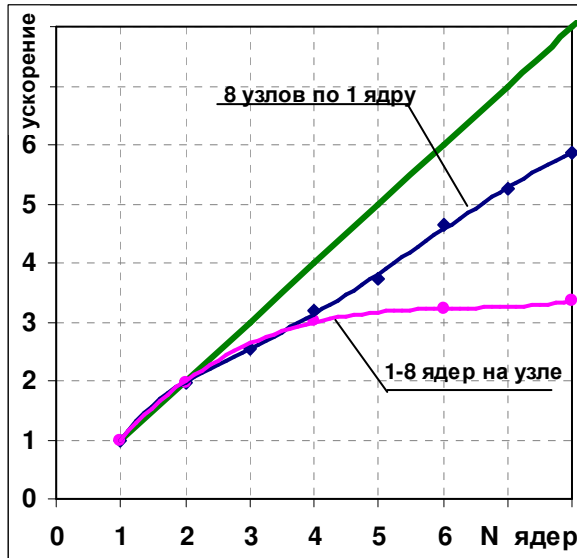


Рис. 8. Масштабируемость малой задачи на узле кластера SKIF Урал при использовании различных способов запуска задачи

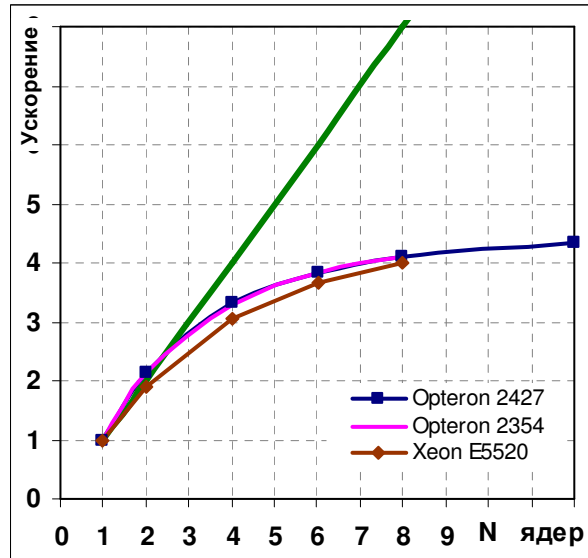


Рис. 9. Масштабируемость малой задачи 2x AMD Opteron 2354 , 2x AMD Opteron 2427, 2 x Intel Xeon E5520

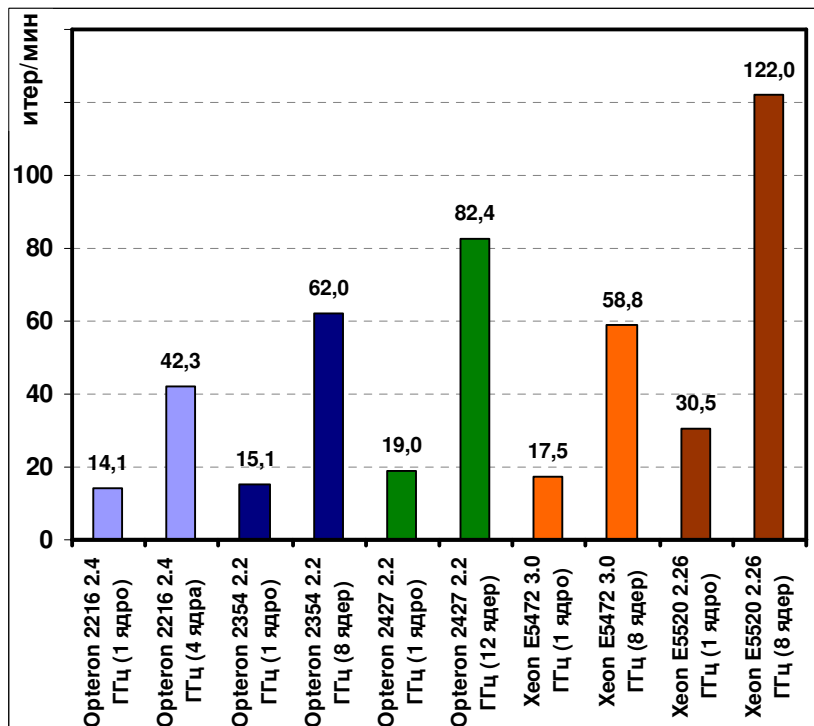


Рис. 10. Абсолютная производительность различных систем при решении малой задачи

Результаты тестирования средней и большой задачи на рабочей станции AMD Opteron 2427 приведены на рис. 10, в тестах были использованы версии решателя с одинарной и двойной точностью. Очевидно, на большой задаче с использованием решателя с одинарной точностью не достигается ограничение пропускной способности шины памяти, кроме того, масштабируемость решателя с двойной точностью превышает масштабируемость на малой и средней задаче (рис. 9, 10). Средняя задача показывает наименьшую масштабируемость среди трех тестовых задач как для решателя с одинарной, так и с двойной точностью. Вероятно, это обуславливается наличием нескольких областей с высоким градиентом скоростей и давления, и, как следствие, существенным различием по сходимости и времени расчета между подобластями декомпозиции расчетной области.

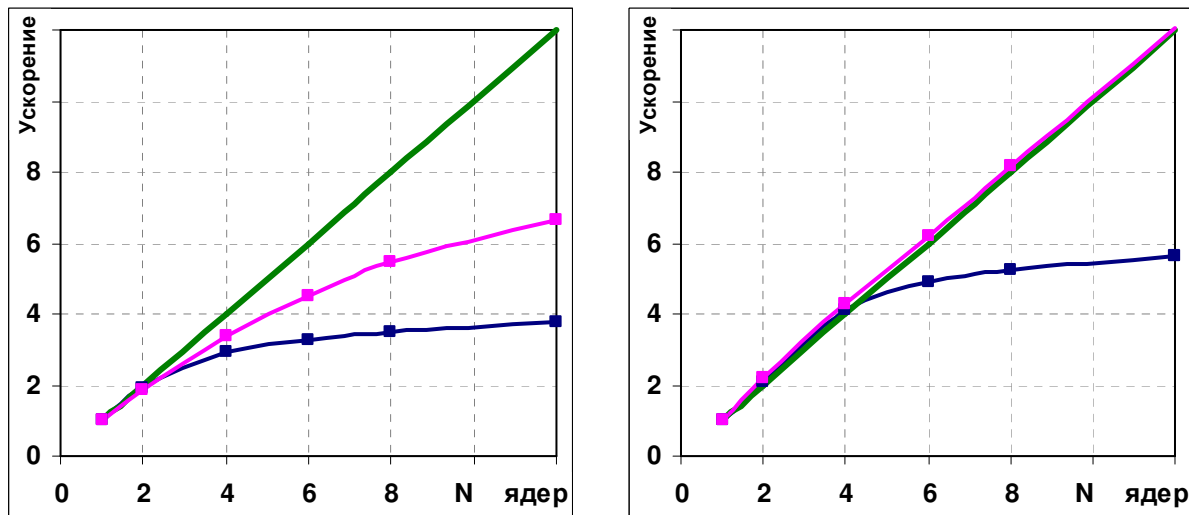


Рис. 10. Масштабируемость средней (слева) и большой (справа) задачи 2x AMD Opteron 2427, решатели одинарной (верхняя линия) и двойной (нижняя линия) точности

Эффективность параллельных вычислений на малых задачах

Эффективное использование кластеров ограничивается размером задачи. Разбиение сетки приводит к созданию областей перекрытия в зонах связи между расчетными подобластями задачи. Процентное отношение количества наложенных (перекрытых) узлов расчетной сетки к общему количеству узлов сетки в оптимальном варианте должно быть минимальным, не более 10 %. Величина перекрытия более 20 % свидетельствует о том, что размер расчетной области на один процесс решателя достиг критического минимума - время вычислений на каждой подобласти будет незначительным по сравнению со временем обмена информацией между ними. Величина перекрытия сохраняется в диагностической информации процесса решателя ANSYS CFX при проведении вычислений.

Тестовые вычисления проводились на малой задаче с числом узлов равным 46 218. Загрузка вычислительных ядер проводилась по самой эффективной схеме – по одному ядру на узел. Эффективное ускорение в 11,3 раза (эффективность 56,3%) было получено при работе кластера на 20 ядрах (рис. 11). При этом увеличение вычислительной мощности происходит практически линейно с коэффициентом ускорения равным 14,3. Минимальное эффективно используемое количество узлов сетки на одно ядро для исследуемой задачи находится в пределах от 2000 до 3000 (около 1500 элементов), что существенно ниже уровня, рекомендуемого документацией ANSYS CFX [1].

Эффективность использования методов декомпозиции расчетной области

Оптимизации процесса распараллеливания сетки для последующего использования кластерных систем происходит с использованием различных типовых программных средств декомпозиции (в терминах ANSYS CFX – "partitioning"). Эффективность этих программных средств в большой мере зависит от типа и геометрии задачи. При проведении параметрических вычис-

лений, требующих больших временных затрат, необходимо выбрать оптимальный продукт, обеспечивающий максимальное ускорение вычислений. Для получения максимального масштабирования рекомендуется использовать предлагаемый по умолчанию декомпозитор (partitioner) "MeTiS". На рис. 7-11 приведены результаты масштабирования, полученные с использованием MeTiS'a. Использование этого программного средства позволило получить эффективное разбиение даже при очень низких числах узлов сетки на одно ядро. При использовании других методов разделения задачи при проведении вычислительных экспериментов было выявлено увеличение процента перекрытия расчетных подобластей по сравнению с MeTiS'ом, и, как следствие значимое снижение скорости вычислений. Эффективность использования MeTiS'a при масштабирования других типов задача требует дополнительных исследований.

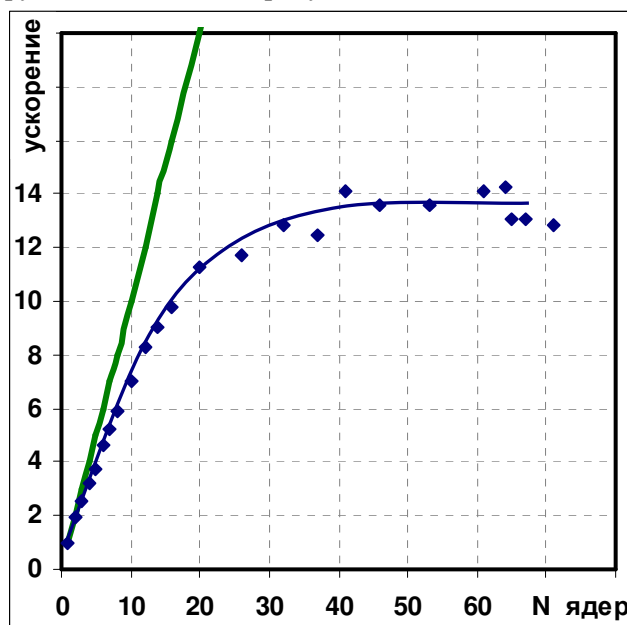


Рис.11 Максимальное масштабирование на малой задаче

Зависимость масштабирования от операционной системы и метода распараллеливания

На рис. 10 приведены результаты расчета малой задачи на двухпроцессорной четырехъядерной рабочей станции, характеристики которой приведены в таблице 2. Расчеты проводились в операционной системе MS Windows XP Professional x64 с использованием метода распараллеливания PVM Local Parallel, задаваемого по умолчанию, с использованием MPICH; в операционной среде Linux с использованием MPICH. Интенсивность вычислений оценивалась по числу итераций в минуту. При использовании метода PVM Local Parallel при увеличении числа вычислительных ядер происходит существенное снижение эффективности работы. При переходе на MPICH увеличение числа процессоров приводит к увеличению скорости вычислений. Максимальное ускорение получено при работе в операционной среде Linux.

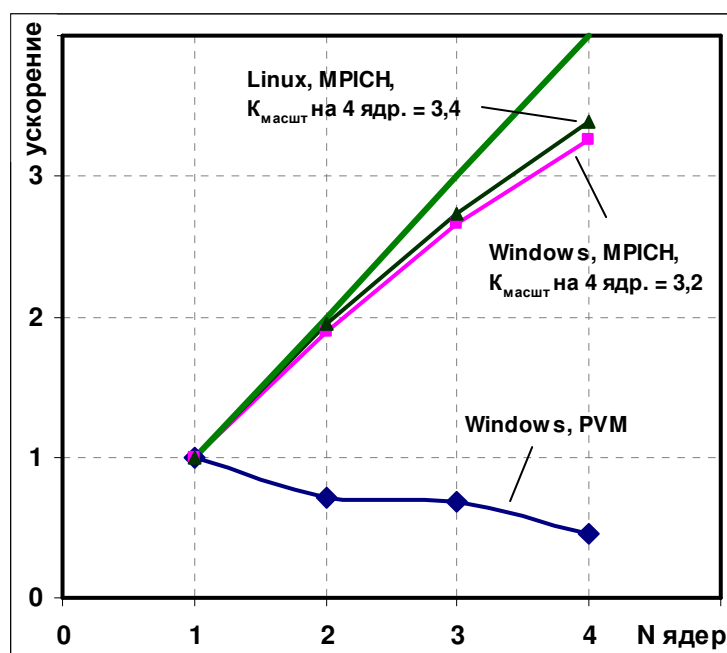


Рис.10. Масштабируемость раб. станции 2x AMD Opteron 2216 в зависимости от операционной системы и метода распараллеливания

Выводы

Вычислительная мощность современных аппаратных платформ и наличие хорошо масштабируемых программных пакетов вычислительной гидроаэродинамики позволяет эффективно решать большие задачи с размером вычислительной сетки в десятки миллионов узлов.

Проведен сравнительный анализ производительности и масштабируемости новейших процессоров AMD Opteron Istanbul и Intel Nehalem

Выявлены проблемы оценки производительности суперкомпьютерных систем – HPL тест не отражает реальное поведение системы на прикладном программном обеспечении (конечно-элементные пакеты, ВГАД)

При наладке кластера или установке нового программного обеспечения необходимо проведение работ по оптимизации загрузки суперкомпьютерной системы различным прикладным программным обеспечением для максимально эффективного использования вычислительных мощностей и имеющихся программных лицензий.

Литература

1. Руководство пользователя ANSYS CFX: ANSYS CFX-Solver Modeling Guide, Using the Solver in Parallel, Advice on Using CFX in Parallel.
2. Интернет-ресурс http://www.fluent.com/software/fluent/fl6bench/fl6bench_12.0/index.htm.
3. А.А.Аксенов, А.А.Дядькин, В.А.Кутин, И.В.Москалёв, Г.Б.Сушко, С.А.Харченко Решение больших задач вычислительной гидродинамики на СКИФ МГУ с помощью FLOWVISION // Материалы Всероссийской научной конференции «Научный сервис в сети ИНТЕРНЕТ: решение больших задач», Новороссийск, 22–27 сентября 2008 г. М.: Изд-во Моск. Уни-та. 2008. С. 69-73.
4. М.А. Еремин, В.Н. Любимов Параллельный код трехмерного моделирования процессов космической газодинамики // Параллельные вычислительные технологии (ПаВТ'2010): Труды международной научной конференции (Уфа, 30 марта - 3 апреля 2010 г.).
5. М.В. Кудрявцев, В.В. Мошкин, М.А. Полуниин Л.К. Эйсымонт Оценочное тестирование кластеров на базе процессоров AMD Barcelona и Shanghai с сетями Infiniband DDR и QDR Вычислительные методы и программирование, 2009, т. 10, стр. 69-77.