

ОТЗЫВ ОФИЦИАЛЬНОГО ОППОНЕНТА

на диссертацию Баранова Антона Викторовича

«Методы и средства управления вычислительными ресурсами в суперкомпьютерных системах коллективного пользования», представленной на соискание ученой степени доктора технических наук по специальности 2.3.5 – Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей

Актуальность темы исследования

Одним из факторов стратегического значения, во многом обеспечивающим технологический суверенитет государства, является уровень развития суперкомпьютерных технологий. Применение суперкомпьютеров как высокопроизводительных вычислительных систем позволяет за счет сокращения сроков научных исследований и разработок промышленных изделий получить конкурентные преимущества в таких областях, как создание новых материалов, энергетика, транспорт, медицина, обеспечение безопасности государства. Суперкомпьютеры являются неотъемлемой частью современной научной инфраструктуры, на их основе реализуется переход к передовым технологиям проектирования и создания высокотехнологичной продукции как один из приоритетов научно-технологического развития Российской Федерации.

Дорогостоящие вычислительные ресурсы суперкомпьютерных систем используются главным образом в режиме коллективного пользования. Пользователи формируют задания, каждое из которых включает расчетную программу, входные данные и требования к ресурсам.

Обработка входного потока пользовательских заданий осуществляется специальными программными системами управления заданиями (СУЗ). В современных суперкомпьютерах на СУЗ возложены основные задачи управления вычислительными ресурсами, а эффективность использования этих ресурсов во многом определяется качеством функционирования СУЗ.

В настоящее время СУЗ представляют собой развитые программные комплексы со сложной архитектурой и широким спектром функциональных возможностей. Одна из важнейших этих функций - планирование, которое заключается в формировании одного или нескольких расписаний запусков заданий. Во многом эффективность той или иной СУЗ определяется качеством планирования заданий. Важность и актуальность создания и развития методов и средств планирования заданий подтверждаются в том числе многочисленными публикациями на эту тему в ведущих отечественных и зарубежных изданиях. Одной из главных проблем в этой области является неоднородность потоков заданий, порождаемая разнообразием пользовательских вычислительных задач и существенным различием в требованиях разных пользователей к суперкомпьютерным ресурсам и времени их использования. Оптимизация планирования заданий дополнительно осложняется тем, что основные показатели качества планирования, такие как загрузка вычислительных ресурсов, среднее время

ожидания задания в очереди, средний коэффициент замедления заданий, являются взаимосвязанными и противоречивыми. При этом улучшение одного из показателей часто приводит к ухудшению значений другого. Актуальным является развитие новых методов и средств планирования заданий.

Не менее важной функцией СУЗ является распределение параллельных процессов пользовательского задания по вычислительным узлам суперкомпьютера. Оптимизация такого распределения призвана минимизировать время, затрачиваемое параллельными процессами задания на информационные обмены. Эта задача решается путем поиска оптимального отображения программного графа задания на граф вычислительных узлов и в общем случае является *NP*-полной. При решении задачи средствами СУЗ поиск отображения должен производиться при каждом запуске задания и за ограниченное время. Актуальным видится создание и развитие эффективных методов и алгоритмов поиска оптимального отображения средствами СУЗ за приемлемое время.

Следующей важной задачей исследования является оптимизация параллельных вычислений с распараллеливанием по данным. При решении вычислительных задач с распараллеливанием по данным часто подготовка заданий и организация параллельных вычислений носят рутинный характер, но в то же время требуют от пользователя достаточно высокой квалификации, поскольку необходимо свести к минимуму накладные расходы на распараллеливание. Разработка методов и средств автоматизации распараллеливания по данным с минимизацией накладных расходов также является актуальной научной задачей.

Перечисленные задачи решаются в рамках создания и развития отечественной системы управления заданиями, что, несомненно, определяет актуальность темы исследования. Несмотря на то, что некоторые известные системы, такие как SLURM или OpenPBS/Torque, свободно распространяются в исходных кодах, их разработка и развитие сосредоточены в западных компаниях. В диссертации отмечена необходимость сохранения компетенций и научно-технологического паритета в этой области и актуальность разработок отечественных систем управления заданиями, обладающих набором функциональных возможностей, соответствующим сложившейся мировой практике.

Структура и содержание диссертации

Диссертация состоит из введения, пяти глав, заключения и списка литературы. Общий объем диссертации – 295 страниц, в том числе 79 рисунков и 22 таблицы. Список литературы состоит из 242 наименований.

Во введении обоснована актуальность работы, сформулированы цель, задачи и научная новизна работы. Представлены основные положения, выносимые на защиту, приведены сведения о примененных методах

исследования, теоретической и практической значимости результатов диссертации и их апробации.

В первой главе представлена архитектура системы управления заданиями. Для построения архитектуры подробно рассмотрены основные задачи управления вычислительными ресурсами в типовой суперкомпьютерной системе коллективного пользования, введено понятие жизненного цикла задания, сформулированы требования к системе управления заданиями. Предложена иерархическая пятиуровневая модель управления вычислительными ресурсами суперкомпьютера. На основе иерархической модели выстроена архитектура отечественной Системы управления прохождением параллельных заданий (СУППЗ), показаны распределение компонентов архитектуры по уровням иерархической модели, а также соответствие предложенной архитектуры требованиям к СУЗ.

Во второй главе рассмотрены предложенные соискателем технические и технологические решения, комплекс которых составил функционал СУППЗ. В главе проведено сравнение СУППЗ с известными системами управления заданиями, такими как SLURM, PBS, LSF, по набору функциональных возможностей, а также сравнение СУППЗ и SLURM по количественным показателям качества: загрузке, среднему времени нахождения заданий в очереди, среднему коэффициенту замедления заданий, доле своевременно обработанных заданий. Результаты сравнительного анализа свидетельствуют о примерном паритете по функционалу и качеству между СУППЗ и ведущими мировыми системами.

В главе приведена обобщенная статистика использования СУППЗ в ведущих отечественных суперкомпьютерах, функционировавших в центрах коллективного пользования МСЦ РАН и ИПМ им. М.В. Келдыша РАН на протяжении 25 лет. За это время услугами СУППЗ воспользовались более 1240 человек из 135 научных и образовательных организаций. Эти пользователи с применением указанных суперкомпьютеров реализовали свыше 530 научно-исследовательских проектов. Такая востребованность СУППЗ свидетельствует о высокой практической значимости результатов диссертации.

В третьей главе рассмотрены предложенные соискателем методы планирования разных классов заданий: заданий с фиксированными параметрами, адаптивных заданий и заданий, поступающих от облачных платформ в виде виртуальных машин и контейнеров.

Метод планирования заданий с фиксированными параметрами основан на разделении всех заданий входного потока на классы ординарных (пакетных), отладочных и фоновых заданий. Главную отличительную особенность предложенного метода определяет наличие фоновых заданий, которые могут выполняться произвольное время, при этом периодически прерываться системой управления с возвращением в очередь. Приведенные в главе результаты анализа статистики суперкомпьютеров под управлением

СУПЗ демонстрируют кратное снижение среднего коэффициента замедления фоновых заданий по сравнению со средним значением этого коэффициента для всего потока заданий. Фактически за счет фоновых заданий обеспечивается эффективность выполнения длительных по времени высокопроизводительных расчетов. При этом предложенный метод позволяет одновременно планировать короткие по времени задания, запускаемые с целью отладки. За счет предложенного автором механизма плавающего резерва вычислительных узлов для отладочных заданий удалось при сохранении высокой загрузки обеспечить низкий коэффициент замедления и высокую долю своевременно обработанных отладочных заданий.

Метод постпланирования адаптивных заданий позволяет заполнять окна в расписании запусков заданий с фиксированными параметрами. Разные размеры этих заданий неизбежно влекут образование таких окон, во время которых простаивает часть вычислительных узлов суперкомпьютера. Адаптивные задания способны изменять требования к времени выполнения и объему вычислительных ресурсов в зависимости от размера обнаруженного в расписании окна. Эффективность метода постпланирования подтверждена результатами имитационного моделирования на входных потоках различной интенсивности. Метод позволяет повысить загрузку суперкомпьютера, при этом другие основные показатели качества – среднее время ожидания заданий в очереди и средний коэффициент замедления – ухудшаются незначительно, если интенсивность основного потока заданий с фиксированными параметрами обеспечивает загрузку не более 0,85.

В главе также рассматриваются нестандартные пользовательские задания, представленные в виде виртуальных машин или контейнеров. Отмечается, что нестандартные задания поступают в суперкомпьютерную систему от облачных платформ. Соискателем предложены два метода планирования таких заданий. Первый метод предполагает, что СУЗ поступает под управление облачной платформы как гипервизор виртуального вычислительного узла с большим числом ядер. Второй метод, наоборот, помещает облачную платформу под управление СУЗ, которая использует инструментарий облачной платформы для разворачивания и запуска нестандартных заданий. В главе представлены варианты реализации предложенных методов для платформ OpenStack и Proxmox.

В четвертой главе рассматривается задача отображения параллельной программы на вычислительные узлы суперкомпьютера. Для ее решения соискателем предложены метод и алгоритмы поиска отображения информационного графа параллельной программы на граф свободных узлов суперкомпьютера.

Метод поиска отображения включает два этапа. На первом этапе осуществляется выбор вычислительных узлов из множества свободных на момент запуска очередного задания. На втором этапе с использованием

выбранных на первом этапе узлов производится поиск оптимального отображения программного графа на граф выбранных узлов. Выполнение первого этапа позволяет сократить размерность задачи отображения на втором этапе и не задерживать при этом очередь заданий. Для обоих этапов предложены эвристические алгоритмы. Первый этап сведен к поиску разрезания графа свободных узлов на минимально связанные между собой подграфы, для чего применен алгоритм, основанный на сочетании имитации отжига и случайного выбора вершин подграфов. Поиск отображения на втором этапе производится параллельным алгоритмом, основанным на циклически повторяющихся фазах имитации отжига и генетического отбора.

В главе приведены результаты исследования характеристик предложенных алгоритмов отображения. Для суперкомпьютера МВС-1000 применение алгоритмов позволило существенно повысить производительность как тестовых программ, так и реальных приложений. Представленные в главе результаты экспериментов по определению точности и скорости поиска отображения параллельным циклическим алгоритмом показали преимущество этого алгоритма перед другими решениями: методами имитации отжига и генетического отбора.

В пятой главе рассмотрены предложенные соискателем метод и программные средства организации параллельных вычислений с распараллеливанием по данным. Основу решений составляет иерархическая система программных менеджеров, осуществляющих рекурсивное разделение входного пула данных на порции для обработки множеством экземпляров прикладной программы. Каждый менеджер выполняется на своем уровне иерархии, управляя либо множеством подчиненных менеджеров нижестоящего уровня, либо экземплярами прикладной программы. Менеджер располагает пулом входных данных, полученным от вышестоящего менеджера. Менеджер разделяет пул на порции заданного размера и асинхронно распределяет эти порции между подчиненными менеджерами или экземплярами программы. Порция (пул) данных представляется в виде упорядоченного набора элементарных порций данных - слайсов. Каждая порция (пул) однозначно определяется номерами первого и последнего слайсов, что определяет в каждом менеджере компактное состояние вычислений в виде остатка входного пула и всех порций, выделенных подчиненным менеджерам. Такое представление вычислений позволяет отказаться от баз учета обработанных порций данных, что в свою очередь позволяет свести к минимуму накладные расходы на распараллеливание.

В главе рассмотрены разработанные автором архитектуры программных комплексов (ПК) «Пирамида» и ХР-СОМ, разработанного в НИВЦ МГУ им. М.В. Ломоносова. ПК «Пирамида» по результатам сравнительных экспериментов показал большие по сравнению с ПК Х-СОМ накладные расходы при обработке входных данных в виде набора строк в файле. В предложенной гибридной архитектуре ПК ХР-СОМ возможности ПК Х-

СОМ совмещены с методом иерархического деления данных. За счет такого совмещения автору удалось добиться меньших накладных расходов на распараллеливание по данным по сравнению с другими известными решениями.

В заключении обобщены основные результаты и выводы диссертации. Приведены сведения о внедрении результатов диссертации в практическую деятельность ряда научных организаций и в учебный процесс подготовки специалистов.

Научная новизна диссертационной работы состоит в следующем.

1. Разработаны методы планирования, совмещающие обработку классов ординарных, отладочных, фоновых и адаптивных заданий, в том числе нестандартных заданий, представленных в виде виртуальных машин и контейнеров.

2. Предложен двухэтапный метод отображения параллельной программы на вычислительные узлы суперкомпьютера: на первом этапе производится выделение вычислительных узлов для параллельной программы и тем самым сокращается размер задачи отображения, на втором этапе выделенные узлы используются для выполнения параллельного алгоритма поиска отображения.

3. Разработан эвристический алгоритм выделения вычислительных узлов для задания, основанный на разрезании графа свободных вычислительных узлов на два минимально связанных друг с другом подграфа при помощи имитации отжига.

4. Разработан параллельный алгоритм поиска отображения программного графа на граф вычислительных узлов с использованием циклической смены фаз имитации отжига и генетического отбора.

5. Предложен метод иерархического деления данных для организации параллельных вычислений с распараллеливанием по данным на основе разделения входного пула данных на наборы упорядоченных элементарных вычислительных работ, обеспечивающий компактное представление состояния вычислений и низкие накладные расходы на распараллеливание.

6. Построена архитектура системы управления заданиями пользователей, основанная на иерархической модели управления вычислительными ресурсами суперкомпьютерной системы коллективного пользования.

Достоверность результатов и выводов диссертации подтверждается:

– 25-летним положительным опытом применения Системы управления прохождением параллельных заданий (СУПЗ) в ведущих российских суперкомпьютерах, установленных в МСЦ РАН, ИТМ им. М.В. Келдыша РАН и ряде других научных и образовательных организаций;

- статистическими данными о работе отечественных суперкомпьютеров, демонстрирующими кратное снижение коэффициента замедления для классов отладочных и фоновых заданий при применении предложенного в диссертации метода планирования заданий с фиксированными параметрами;

- результатами имитационного моделирования, продемонстрировавшими для предложенного метода постпланирования максимизацию загрузки суперкомпьютера при незначительном увеличении коэффициента замедления основного потока заданий с фиксированными параметрами;

- результатами экспериментов на суперкомпьютерах серии МВС по применению предложенных в диссертации метода и алгоритмов отображения параллельной программы на вычислительные узлы суперкомпьютера;

- результатами экспериментального сравнения с известными решениями (MapReduce, BOINC, X-COM) программных комплексов «Пирамида» и XP-COM, в которых реализован метод иерархического разделения данных для организации параллельных вычислений с распараллеливанием по данным;

- всесторонней апробацией основных результатов диссертации на ряде международных и всероссийских научных конференций;

- 45 печатными работами по теме диссертации, опубликованными в ведущих рецензируемых отечественных и зарубежных изданиях.

Степень обоснованности научных положений, выводов и рекомендаций, сформулированных в диссертации

Проведенное в работе сравнение методов, алгоритмов и программных средств, составивших научные положения, выводы и рекомендации диссертации, показало соответствие предложенных соискателем архитектурных, технических и технологических решений мировому уровню. Ход и результаты проведенных автором исследований в диссертации изложены последовательно и логично, предложенные решения убедительно аргументированы. Выводы по главам и положения заключения адекватно резюмируют представленный в тексте материал.

Вынесенные на защиту научные положения, а также сформулированные в диссертации выводы и рекомендации обоснованы соискателем в достаточной степени.

Ценность для науки и практики результатов работы

Практическая значимость результатов работы подтверждается их успешным использованием в ряде реальных научных проектов и экспериментальных исследований.

Представленные в диссертации архитектурные, технические и технологические решения реализованы в составе Системы управления прохождением параллельных заданий (СУППЗ), нашедшей практическое применение в отечественных суперкомпьютерах МВС-1000, МВС-1000/16, МВС-1000/32, МВС-1000М, МВС-15000ВМ, МВС-6000ИМ, МВС-100К, МВС-10П, МВС-Экспресс, К-100, К-60. Указанные системы обеспечивают высокопроизводительные расчеты в суперкомпьютерных центрах коллективного пользования МСЦ РАН – НИЦ «Курчатовский институт», ИПМ им. М.В. Келдыша РАН для более чем 1240 пользователей-исследователей из 135 научных и образовательных организаций.

Разработанные эвристические алгоритмы разрезания графа свободных узлов и поиска отображения программного графа на граф вычислительных узлов суперкомпьютера вносят вклад в решение фундаментальной квадратичной задачи о назначениях. Теоретические положения и накопленный практический опыт, представленные в диссертации, могут служить основой дальнейшего развития методов и средств построения систем управления вычислительными ресурсами и организации параллельных вычислений в суперкомпьютерных центрах.

Подтверждение опубликования основных результатов диссертации в научной печати

Описание каждого полученного результата в тексте диссертации и автореферате снабжено библиографическими ссылками на соответствующие публикации соискателя, что позволяет подтвердить опубликование и апробацию основных результатов диссертации в научной печати. Всего по теме диссертации автором опубликовано 45 печатных работ, из них 27 работ опубликованы в журналах, рекомендованных ВАК, в том числе 20 работ – в журналах, входящих в индекс RSCI, 4 работы – в журналах, входящих в список K1 перечня ВАК, 15 работ – в журналах, индексируемых в международных базах научных изданий Web of Science и Scopus. По теме диссертации получено 6 свидетельств о государственной регистрации программ для ЭВМ и баз данных.

Соответствие содержания автореферата основным положениям диссертации

В автореферате приведены сведения об актуальности исследования, цели и задачах диссертационной работы, научной новизне, апробации, теоретической и практической значимости результатов диссертации, о выносимых на защиту положениях. В тексте автореферата полно раскрыто основное содержание глав диссертации, приведены основные публикации по теме исследования. Автореферат диссертации адекватно отражает ее содержание.

Замечания по работе

1. В подразделе 3.1.2, посвященном планированию адаптивных заданий, упоминается экосистема PanDA, изначально ориентированная на эксперимент ATLAS (Большой адронный коллайдер). Здесь было бы уместно привести данные и по платформе DIRAC, примененной в эксперименте LHCb. Сравнительный анализ DIRAC и PanDA, в том числе систем управления потоками задач, придал бы больший вес материалу по методам и средствам планирования адаптивных заданий, в частности, на основе грид-технологий и облачных платформ.

2. В разделах 3.1 и 3.5, в которых рассматриваются методы и средства планирования стандартных и нестандартных заданий в облачных вычислениях, среди основных моделей обслуживания выделяются программное обеспечение как услуга (SaaS), платформа как услуга (PaaS) и инфраструктура как услуга (IaaS). Тем не менее этими моделями далеко не все исчерпывается в области планирования и управления ресурсами в облачных вычислениях. Так, в настоящее время активно продолжает развиваться парадигма «поток работ как сервис» (Workflow as a Service - WaaS) для выполнения наукоемких приложений в таких областях как сейсмология (CyberShake), биоинформатика (Epigenomics, SIPHT), астрофизика (Montage), физика гравитационных волн (LIGO), гидро и аэродинамика, квантовая химия, нанотехнологии, гидрометеорология, моделирование социальных систем и транспортной инфраструктуры. Платформы WaaS строятся, как правило, на расширении функционала систем управления потоками работ, таких как ASKALON, Galaxy, HyperFlow, Kepler, Pegasus, Taverna, CloudBus и ряда других, в том числе с открытыми кодами и применимыми в разных предметных областях. Изложение разделов 3.1 и 3.5 выиграло бы при соответствующем анализе упомянутых проектов.

3. Следовало провести более глубокий анализ опыта успешной интеграции облачных и высокопроизводительных вычислений в России. В частности, в настоящее время в Объединённом институте ядерных исследований (ОИЯИ) активно развивается проект ускорителя NICA с экспериментами BM@N, MPD, SPD. Для решения задач обработки данных в этих экспериментах задействуется широкий спектр вычислительных ресурсов: кластеры Tier1/Tier2, суперкомпьютер «Говорун», облачные инфраструктуры ОИЯИ и стран-участниц, локальные кластеры NICA, DCC, а также системы дискового и ленточного хранения.

Заключение по работе

Отмеченные выше замечания, в целом, не снижают высокого научного уровня и практической значимости работы. Замечания не являются критичными и не уменьшают важности диссертационной работы, в которой изложены востребованные практикой технические и технологические решения по созданию и развитию методов и средств управления

вычислительными ресурсами суперкомпьютерных систем коллективного пользования.

Диссертация обладает внутренним единством, представляет законченную научно-квалификационную работу, выполненную на актуальную тему, и содержит новые научно обоснованные архитектурные, технические и технологические решения, внедрение которых вносит значительный вклад в развитие научных суперкомпьютерных центров коллективного пользования как неотъемлемой части исследовательской инфраструктуры страны.

Считаю, что диссертационная работа «Методы и средства управления вычислительными ресурсами в суперкомпьютерных системах коллективного пользования» удовлетворяет требованиям, предъявляемым к диссертациям по специальности 2.3.5 – Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей, а ее автор Баранов Антон Викторович заслуживает присуждения ему ученой степени доктора технических наук по этой специальности.

Официальный оппонент

доктор технических наук (2.3.2 – Вычислительные системы и их элементы), профессор, заведующий кафедрой вычислительных технологий Национального исследовательского университета «МЭИ»,

 Виктор Васильевич Топорков

«24» нояб 2025 г.

ПРИСТАВЛЯЮ

ПОДПИСЬ НАЧАЛЬНИКА УПРАВЛЕНИЯ
ПЕРСОНАЛОМ Л.И.П. ВАЯ

Сведения об организации
Федеральное государственное
высшее образование
«МЭИ»



национальное образовательное учреждение
исследовательский университет

111250, г. Москва, вн. тер. г. муниципальный округ Лефортово, ул.
Красноказарменная, д. 14, стр. 1.; +7 495 362-75-60; электронная почта:
universe@mpei.ac.ru; адрес в сети Интернет: <https://mpei.ru/Pages/default.aspx>