

СУПЕРКОМПЬЮТИНГ НА ПРЕДПРИЯТИИ. ВЫБОРЫ И РЕШЕНИЯ

Петунин Сергей Александрович, к.ф.-м.н., Новиков Алексей Борисович, Иванов
Константин Владимирович

ФГУП «Всероссийский научно-исследовательский институт автоматики»
им.Н.Л.Духова (г.Москва)

В докладе описана практика внедрения суперкомпьютерных технологий в отраслевом НИИ, базирующаяся на сочетании использования локальных и удаленных суперкомпьютеров с open-source платформами математического моделирования и на применении комплексного подхода к автоматизации сопровождения расчетов. Представлены разработки системного программного обеспечения, проводимые в рамках выбранных решений.

Ключевые слова: вычислительный кластер, сеть доступа, автоматизированная система, планирование заданий, система мониторинга

1. Введение

В связи с организацией на базе ФГУП «ВНИИА» Центра Фундаментальных и Прикладных Исследований (ЦФПИ) ГК «Росатом» в институте возникла необходимость массового применения аппарата математического моделирования и, как следствие, проведения суперкомпьютерных расчетов.

Для любого предприятия с развитой ИТ-инфраструктурой при появлении задач, требующих суперкомпьютерных мощностей, встают различные проблемы системного уровня. В их числе следует перечислить проблемы выбора высокопроизводительных аппаратно-программных платформ, распределения долей между разработкой и внедрением, а также необходимость «мягкого» включения супер-ЭВМ в существующую корпоративную технологию обработки информации.

2. Технические решения

Полученный на данный момент опыт можно кратко сформулировать в следующих тезисах:

- 2.1 Аппаратная база проведения расчетов построена на двух локальных кластерах: технологическом (ТК) и расчетном (МВК-Т). На технологическом кластере производится компиляции, сборка и отладка тестовых конфигураций прикладного программного обеспечения и осуществляются предварительные расчеты. Идентичность программных сред ТК и МВК-Т позволяет переносить отлаженные сборки на расчетный кластер, тем самым повышая надежность его функционирования.
- 2.2 Второй уровень расчетов обеспечивается организацией удаленного доступа к суперкомпьютерным центрам коллективного пользования МГУ[1] и ФГУП «РФЯЦ-ВНИИЭФ»[2]. При этом наблюдаются сильные различия в профилях потребленных ресурсов у рассчитанных заданий: на локальном уровне осуществляется большое количество запусков задач со сдвигом в сторону малопроцессорных геометрий и коротких времен, а на удаленных кластерах, наоборот, характерно малое количество запусков с существенным запросом на вычислительные и временные ресурсы (рис.1).

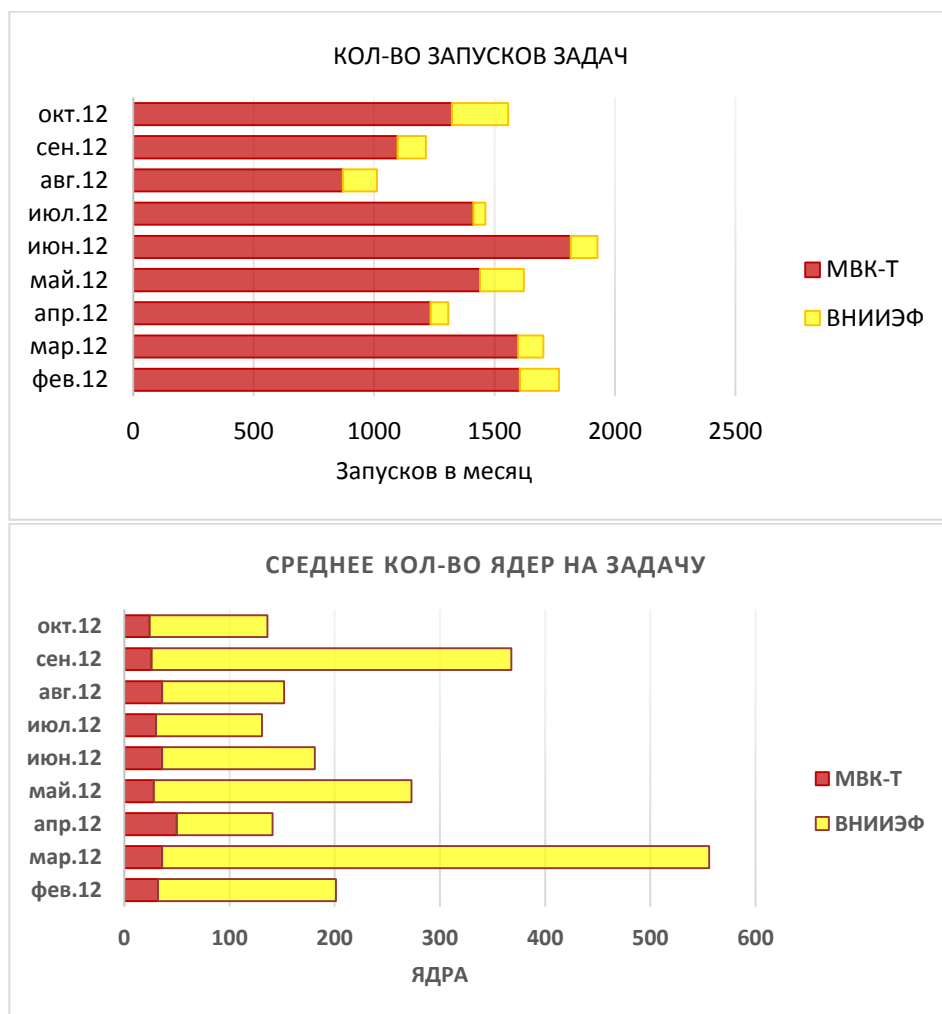


Рис.1 Профили заданий локального и удаленного счета

- 2.3 Ориентация пользователей ЦФПИ на использование широкого набора прикладных моделирующих пакетов, доступных для установки по свободным лицензиям (Lammps, Abinit, Firefly, Geant4 и др.), позволило уже в первый месяц после запуска MBK-T в эксплуатацию получить 60% загрузки вычислительного поля. За два года средний показатель загрузки только вырос.
- 2.4 Построение сети доступа к кластерным ресурсам на инфраструктуре корпоративной сети института решило проблему подключения рабочих мест пользователей. Однако, на данный момент к нерешенным задачам относятся, в частности, интеграция расчетных технологий с используемыми в сети технологическими сервисами, а также проблема обеспечения защиты информации в условиях постоянно изменяемой программной среды разработчика, параллельной архитектуры кластера и поддержки групп пользователей с различными правами доступа.

3. Разработка автоматизированной системы управления кластерами

В условиях количественного роста вычислительных кластеров в институте и качественного усложнения их архитектур целесообразно создание эффективной и надежной системы управления, базирующейся на оценке состояния кластерных

ресурсов и предоставляющей единую технологию обработки. Такую систему невозможно спроектировать только на базе разработки системного программного обеспечения кластера, расширяющего возможности его операционной системы. Подходящей концепцией, иллюстрирующей организационно-программную модель этой системы, является ее определение как автоматизированной системы (АС), т.е. системы, включающей в себя объекты оборудования, сети доступа, системные процессы, прикладные программы, интерфейсы и т.д., а также субъекты–пользователей. В реализации функционала АС можно выделить две приоритетные задачи.

3.1 Задача 1. Разработка адаптивной подсистемы управления вычислительными заданиями с высокой утилизацией кластерных ресурсов [3].

Эксплуатация высокопроизводительных вычислительных кластеров в общем случае осуществляется в двух режимах: коллективном и корпоративном. В первом случае кластеры являются вычислительной базой крупных центров коллективного пользования с массовым и случайным характером формирования вычислительных заданий. Во втором случае, суперкомпьютер, как правило, используется в рамках решения задач одного или нескольких предприятий с определенным набором прикладных программ. При корпоративном режиме становится возможной адаптация принципа управления заданиями с учетом особенностей производственного счета. Разработка планировщика заданий, связанного с историей выполнения прикладных программ и обладающего свойствами многовариантного динамического подбора подходящего алгоритма планирования, позволит обеспечить высокую эффективность аллокации и использования вычислительных и коммуникационных ресурсов. В настоящее время завершено создание прототипа данной системы (рис.2).

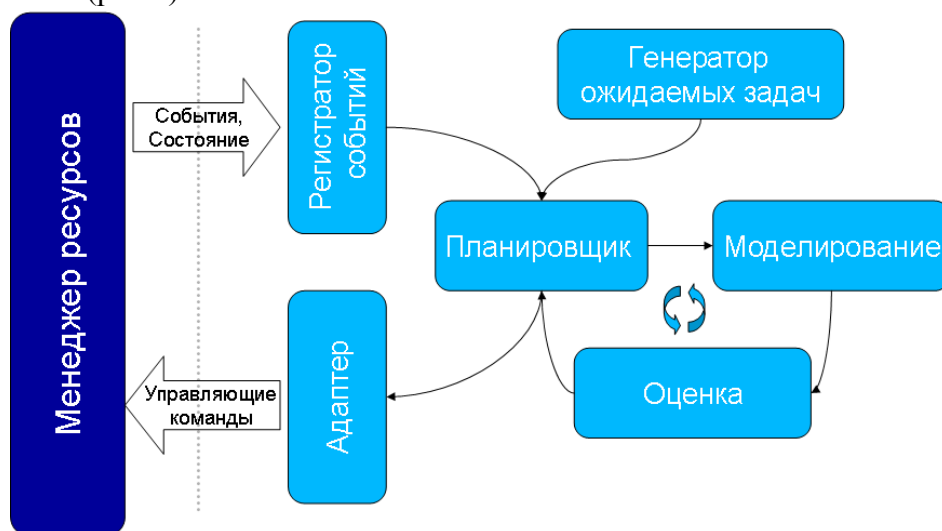


Рис.2 Архитектура системы управления заданиями для АС

3.2 Задача 2. Архитектура мониторинговой подсистемы с прогнозированием ошибок.

Актуальность разработки такой подсистемы вызвана необходимостью увеличения доли автоматического «интеллекта» при анализе мониторинговых данных суперкомпьютерных расчетов, который только появляется в аналогичных проектах, таких как Clustrx Watch (разработка Т-Платформс) или киевская система SCMS [4]. Подсистема спроектирована с учетом оценки состояния как аппаратных и системных параметров, так и базы знаний по прикладным расчетам, и включает в

себя обратную связь с уровнями инженерии, вычислительного поля, коммуникационной системы, системы хранения данных, системы управления заданиями. Модуль сбора данных реализован как безагентная компонента, основанная на использовании SSH-туннелирования для доступа к узлам и получения данных через системные интерфейсы (виртуальную файловую систему, службы уровня ядра и др.). Оптимизация хранения данных достигается за счёт интервального метода снятия данных, их аккумуляции и резервного копирования. В настоящее время реализуется первая версия мониторинговой подсистемы с модулями анализа, обратной связи, визуализации и шаблоном модуля прогнозирования. Разработка ведется в многоязыковой среде (C, Python + JavaScript, Qt C++). Прототип системы используется на производственном расчетном кластере MBK-T.

Литература

1. Воеводин Вл.В. и др. «Суперкомпьютерный комплекс МГУ: архитектура, пользователи, задачи» // МСКФ-3, презентация доклада, 2012.
2. Шагалиев М.Р. и др. «Актуальные проблемы создания и внедрения технологий суперкомпьютерного моделирования в науку и промышленность» // МСКФ-3, презентация доклада, 2012.
3. Новиков А.Б., Петунин С.А. «Влияние специализированных алгоритмов планирования заданий на эффективность использования вычислительных ресурсов в частных случаях» // XIII международный семинар «Супервычисления и математическое моделирование», тезисы доклада, РФЯЦ-ВНИИЭФ, 2011.
4. Головинский А.Л. и др. «Система управления суперкомпьютером SCMS» // Сборник трудов ПАВТ'12, 2012.