

Моделирование потока задач вычислительного кластера НИУ ВШЭ с использованием SLURM Simulator

Р.М. Мишенин, П.С. Костенецкий

Высшая школа экономики

Задача эффективного распределения ресурсов вычислительной системы широко известна. Она становится критически важной в многопользовательских многопроцессорных системах, таких как суперкомпьютеры. В НИУ ВШЭ функционирует высокопроизводительный вычислительный кластер «сHARISMa», состоящий из 48 вычислительных узлов шести типов. Основными характеристиками узлов являются наличие графических ускорителей и их модели, типы центральных процессоров и объем оперативной памяти. На суперкомпьютере используется планировщик задач SLURM версии 23.02.4.

При подборе параметров планировщика необходимо заранее определить, как они повлияют на поток задач вычислительного кластера. Одним из способов решения данной проблемы является моделирование вычислительного кластера с учетом его топологии и многообразия выполняемых задач. В настоящее время существует множество симуляторов высокопроизводительных кластеров, таких как AccaSim, GridSim, CloudSim, SimGrid. В результате сравнения было принято решение использовать SLURM Simulator [1].

Для проведения моделирования был подготовлен комплект конфигурационных файлов и трасс событий. В набор входят конфигурационные файлы SLURM, файл настроек менеджера учетных записей, файл, содержащий перечень аккаунтов и пользователей, зарегистрированных в системе управления заданиями. Каждое событие содержит временную метку постановки задания в очередь, количество запрошенных ресурсов, класс обслуживания и другие стандартные атрибуты задачи, необходимые для дальнейшего анализа и моделирования. В качестве источника данных для формирования файла с трассами событий был использован поток задач, выполненных за год на вычислительном кластере «сHARISMa».

Под модельным временем выполнения потока задач будем понимать время, необходимое для выполнения заданного набора задач на моделируемой конфигурации вычислительного комплекса. Задача сводится к минимизации модельного времени путем оптимизации конфигурации моделируемого суперкомпьютера. Ключевыми параметрами, существенно влияющими на работу алгоритма планирования, являются тип планировщика (builtin или backfill), а также весовые коэффициенты, варьирующиеся от 0 до 2 147 483 645 и определяющие приоритет задач. При помощи моделирования будет подобран оптимальный комплект вычислительного оборудования для апгрейда суперкомпьютера в 2026 г., который наиболее эффективно снизит время ожидания реальных вычислительных задач в очереди.

В качестве проверки модели было рассчитано влияние небольшого апгрейда, выполненного в 2024 г. В кластер были добавлены два вычислительные узла на базе GPU Nvidia H100. Среднее модельное время ожидания задач в очереди уменьшилось на 2.1 % относительно базовой конфигурации, что подтвердилось реальными статистическими значениями, полученных на суперкомпьютере – 1.95 %.

Дальнейшими задачами исследований является подстройка алгоритмов планировщика задач на моделируемой копии суперкомпьютера «сHARISMa». В результате моделирования будут определены изменения в параметрах конфигурации SLURM, внедрение которых в планировщик реального вычислительного кластера позволит снизить среднее время ожидания задач в очереди и дополнительно повысить эффективность использования суперкомпьютера.

Литература

1. Simakov N. et al. Slurm Simulator: Implementation and Parametric Analysis // High Performance Computing Systems. Performance Modeling, Benchmarking, and Simulation. PMBS 2017. Vol. 10724 / eds. by S. Jarvis, S. Wright, S. Hammond. Springer, Cham, 2017. P. 197–217. Lecture Notes in Computer Science. DOI: 10.1007/978-3-319-72971-8_10.