

Двухфазное планирование выполнения композитных приложений в облачных средах

МАРЬИН СЕРГЕЙ

Санкт-Петербургский государственный университет информационных технологий, механики и оптики
e-mail: sergey.maryin@gmail.com

Повышение эффективности использования облачных вычислительных ресурсов является актуальной проблемой в свете стремительного роста популярности общественных облаков. Имеющиеся в этих облаках механизмы автоматического масштабирования весьма примитивны [1], а задача планирования, т.е. распределения по вычислительным ресурсам множества задач, связанных по данным, составляющих единое композитное приложение (workflow), является NP-трудной. Существующие подходы [2] не учитывают в достаточной мере фундаментальной ориентированности облачных вычислений на конечного пользователя. По сути, пользователя интересует в первую очередь цена арендуемых в облаке ресурсов, а не распределение задач по виртуальным машинам, если задачи успевают выполниться в заданный срок. Такая точка зрения позволяет рассматривать процесс планирования разделённым на две фазы: оптимизацию выбора вычислительных ресурсов и распределение задач по выбранным ресурсам. Представленное в данной работе разделение процесса планирования позволяет подбирать алгоритм для каждой из фаз независимо - выходные данные алгоритма оптимизации ресурсов (набор используемых ресурсов) будут входными данными для алгоритма распределения задач. Эффективность планирования повышается за счёт дополнительных итераций фазы оптимизации ресурсов, использующей результаты планирования задач как оценку качества выбора ресурсов. В фазе распределения задач могут использоваться как существующие алгоритмы планирования для облаков, так и алгоритмы, разработанные без учёта специфики облачных вычислений. Для проверки эффективности разработанного подхода было произведено имитационное моделирование. Двухфазное планирование осуществлялось на основе генетического алгоритма, работающего поверх алгоритма облачного планирования IC-PCPD2 [3]. Сравнение эффективности с этим же (IC-PCPD2) алгоритмом, работающим в штатном режиме, показало уменьшение стоимости выполнения пользовательского композитного приложения в среднем на 49%.

Работа выполнена при поддержке гранта Президента Российской Федерации для государственной поддержки молодых российских учёных № МК-6628.2014.9.

Список литературы

- [1] MAO M., HUMPHREY M. Auto-scaling to minimize cost and meet application deadlines in cloud workflows // Int. Conference High Performance Computing, Networking, Storage and Analysis. — 2011. — С. 1–12.
- [2] SINGH L., SINGH S. A Survey of Workflow Scheduling Algorithms and Research Issues // International Journal of Computer Applications. — 2013. — Т. 74, № 15, С. 21-28.

- [3] ABRISHAMI S., NAGHIBZADEH M., ЕРЕМА D. H. J. Deadline-constrained workflow scheduling algorithms for Infrastructure as a Service Clouds // Future Generation Computer Systems. — 2013. — Т. 29, № 1, С. 158–169.