

Применение биоинспирированных методов для решения задачи распределения ресурсов в облачных платформах

Д.И. Парфёнов¹, И.П. Болодурина¹, Л.С. Забродина¹, Л.Ю. Кузнецова¹, Н.А. Янишевская¹

¹Оренбургский государственный университет, проспект Победы 13, Оренбург, Россия, 460018

Аннотация

В данной статье рассматривается применение биоинспирированных алгоритмов к задаче оптимизации ресурсов при обработке больших данных с использованием облачных платформ. Особенностью исследования является учет соглашений об уровне предоставления услуг, являющихся критически важным фактором ресурсного обеспечения. Установлено, что наиболее эффективными методами для решения данной задачи являются гравитационный поиск и алгоритм муравьиной колонии.

Ключевые слова

Облачная платформа, SLA, распределение ресурсов, большие данные

1. Введение

В настоящее время для хранения и обработки больших данных чаще всего используются ресурсы облачных платформ. Одним из главных достоинств распределенных облачных ресурсов является гибкое масштабирование. Таким образом, наиболее эффективным инструментом оптимизации затрат при обработке больших данных является качественно спроектированный механизм, осуществляющий оптимизацию потока данных с учетом SLA.

2. Основная часть

Современный подход к обработке больших данных должен учитывать механизмы эффективного распределения ресурсов. Задачей алгоритмов планирования является гарантированное удовлетворение требований приложений и пользователей согласно SLA.

Рассмотрим задачу размещения виртуальных машин (ВМ) внутри облачной платформы. Найдем баланс между количеством задействованных серверов и количеством ресурсов, требуемых для оптимального распределения задач, путем минимизации функционала:

$$f = \alpha \sum_{i=1}^k s_i + (1 - \alpha) \left[\sum_{i=1}^k \sum_{j=1}^n time_j^{dis} \cdot (v_{ij}^l \oplus v_{ij}) + \sum_{i=1}^k \sum_{j=1}^n time_j v_{ij} \right]. \quad (1)$$

При следующих ограничениях:

$$\sum_{i=1}^k v_{ij} = 1, \quad \forall j \in \overline{1, n}, \quad (2)$$

$$\sum_{j=1}^q v_{ij} \cdot \sum_{x=1}^{num} c_{jx} \leq s_i \cdot \sum_{x=1}^{num} C_{ix}, \quad \forall i \in \overline{1, k}. \quad (3)$$

Здесь k – количество серверов; v_{ij}^l – аналогично v_{ij} , но на предыдущем распределении; n – количество задач (ВМ); $time_j^{dis}$ – время перераспределения задач; q – номер текущей задачи; c_j – ресурсы для решения j -ой задачи; C_i – ресурсы i -го сервера; $num = 3$ – количество типов ресурсов $\{RAM, CPU, Storage\}$; α – коэффициент значимости; $time_j$ – время выполнения j -ой задачи; $s_i = \begin{cases} 1, & \text{если } i\text{-й сервер используется,} \\ 0, & \text{иначе} \end{cases}$; $v_{ij} = \begin{cases} 1, & \text{если } j\text{-я ВМ на } i\text{-м сервере,} \\ 0, & \text{иначе} \end{cases}$.

Ограничение (2) означает, что ВМ должна сопоставляться одному серверу. Неравенство (3) описывает ограничение ресурсов. Таким образом, получили задачу линейного программирования.

В рамках исследования проведен сравнительный анализ интеллектуальных алгоритмов: Genetic Algorithm (GA), Ant Colony Algorithm (ACA), Particle Swarm Optimization (PSO) и Gravitation Search (GS) для решения задачи оптимального распределения ресурсов. Моделирование процессов распределения ресурсов проводилось в облачном центре обработки данных, где планировалось выполнение 5 вариантов задач с различным числом заявок на ВМ.

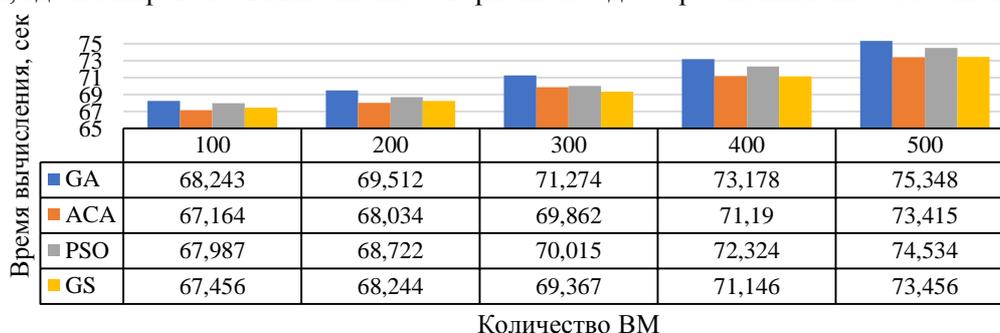


Рисунок 1: Зависимость времени перераспределения ресурсов от количества ВМ

Согласно полученным результатам, наименьшее время на перераспределение ВМ среди имеющихся серверов получено с помощью интеллектуальных методов GS и ACA, а также по мере увеличения количества ВМ данные алгоритмы показали наименьшее число используемых серверов. Отметим, что прогнозируемая точность рассматриваемых методов в разрезе объемов используемых ресурсов позволяет эффективно осуществлять планирование и распределять нагрузку между серверами, а доступные ресурсы не простаивают и используются оптимально.

3. Заключение

В рамках данной работы реализованы четыре различных биоинспирированных метода (GS, ACA, PSO, GA) для решения задачи оптимизации затрат при обработке больших данных в облачных инфраструктурах с учетом соглашений об уровне предоставления услуг. Установлено, что методы GS и ACA позволяют наиболее эффективно использовать доступные параметры серверов в сравнении с другими исследуемыми интеллектуальными методами.

4. Благодарности

Исследование выполнено при финансовой поддержке РФФИ (проект 20-07-53019) и гранта Президента Российской Федерации для государственной поддержке ведущих научных школ Российской Федерации (НШ-2502.2020.9), а также стипендии Президента Российской Федерации молодым ученым и аспирантам (СП-3652.2021.5).

5. Литература

- [1] Krishnan, P. Self-adaptive PSO memetic algorithm for multi objective workflow scheduling in hybrid cloud / P. Krishnan, J. Aravindhar // Int Arab J Inf Technol. – 2019. – Vol. 16. – P. 928-935.
- [2] Shooli, R.G. Using gravitational search algorithm enhanced by fuzzy for resource allocation in cloud computing environments / R.G. Shooli, M.M. Javidi // SN Applied Sciences. – 2020. – Vol. 2(195). – P. 13.
- [3] Legashev, L.V. An Effective Scheduling Method in the Cloud System of Collective Access, for Virtual Working Environments / L.V. Legashev, I.P. Bolodurina // Acta Polytechnica Hungarica. – 2020. – Vol. 17(8). – P. 179-195.