

УДК 004.89

**ИНТЕЛЛЕКТУАЛЬНЫЕ МЕТОДЫ И АЛГОРИТМЫ ПРИНЯТИЯ
РЕШЕНИЙ ПРИ РАСПРЕДЕЛЕНИИ ВЫЧИСЛИТЕЛЬНЫХ РЕСУРСОВ
В ОБЛАЧНЫХ СРЕДАХ**

Волков Захар Иванович

*магистрант, кафедра прикладной информатики и информационных
технологий,*

*Белгородский государственный национальный исследовательский
университет,*

РФ, г. Белгород

**INTELLIGENT METHODS AND ALGORITHMS FOR ADAPTIVE
RESOURCE ALLOCATION IN CLOUD COMPUTING ENVIRONMENTS**

Volkov Zahar Ivanovich

Master's Student, Department of Applied Informatics and Information Technologies,

Belgorod State National Research University,

Russia, Belgorod

АННОТАЦИЯ

В статье рассматриваются современные подходы к разработке интеллектуальных методов и алгоритмов принятия решений при распределении вычислительных ресурсов в облачных средах. Особое внимание уделяется адаптивным, эвристическим и основанным на машинном обучении механизмам управления ресурсами, обеспечивающим баланс между производительностью, стоимостью и качеством обслуживания (QoS). Анализируются проблемы динамичности нагрузки, гетерогенности инфраструктуры и неопределённости входных данных. Предлагается обобщённая архитектура системы поддержки принятия решений для облачной инфраструктуры и описываются перспективные направления исследований.

ABSTRACT

The article examines modern approaches to the development of intelligent methods and decision-making algorithms for resource allocation in cloud computing environments. Particular attention is paid to adaptive, heuristic, and machine learning-based resource management mechanisms that ensure a balance between performance, cost, and Quality of Service (QoS). The challenges of dynamic workloads, infrastructure heterogeneity, and input uncertainty are analyzed. A generalized architecture of a decision support system for cloud infrastructure management is proposed, and перспективные research directions are outlined.

Ключевые слова: облачные вычисления, распределение ресурсов, принятие решений, алгоритмы планирования, виртуализация, машинное обучение, оптимизация.

Keywords: cloud computing, resource allocation, decision-making, scheduling algorithms, virtualization, machine learning, optimization.

Облачные вычисления в настоящее время являются ключевым элементом цифровой трансформации экономики и науки. Крупнейшие провайдеры, такие как Amazon Web Services, Microsoft Azure и Google Cloud Platform, предоставляют масштабируемые вычислительные мощности, позволяющие организациям гибко управлять ИТ-инфраструктурой без необходимости поддержания собственных центров обработки данных. В основе эффективного функционирования таких систем лежит задача рационального распределения вычислительных ресурсов между множеством пользователей и приложений.

Распределение ресурсов в облачной среде представляет собой сложную многокритериальную задачу оптимизации. Необходимо учитывать производительность приложений, соблюдение соглашений об уровне обслуживания, минимизацию затрат и энергопотребления, а также устойчивость системы к пиковым нагрузкам. Дополнительную сложность создают динамичность пользовательских запросов и гетерогенность инфраструктуры, включающей различные типы виртуальных машин, контейнеров и физических серверов.

Традиционные алгоритмы планирования, такие как Round Robin, First Fit, Best Fit, Min-Min и Max-Min, обеспечивают базовую функциональность распределения задач. Они характеризуются простотой реализации и невысокой вычислительной сложностью, однако их эффективность снижается в условиях быстро меняющейся нагрузки. Такие методы не способны в полной мере учитывать многокритериальный характер задачи и адаптироваться к изменяющимся условиям среды.

Для повышения качества распределения ресурсов применяются эвристические и метаэвристические методы, включая генетические алгоритмы, алгоритмы роя частиц и имитацию отжига. Эти подходы позволяют находить приближённо оптимальные решения в сложных пространствах поиска и учитывать несколько критериев одновременно. Однако их применение требует дополнительных вычислительных затрат и тщательной настройки параметров.

Современное направление развития связано с использованием методов машинного обучения. В частности, обучение с подкреплением позволяет формировать адаптивные стратегии управления ресурсами, основанные на накопленном опыте взаимодействия системы со средой. Нейронные сети применяются для прогнозирования нагрузки и выявления скрытых закономерностей в поведении пользователей. Использование предиктивных моделей даёт возможность заблаговременно масштабировать ресурсы и предотвращать нарушения SLA.

Эффективная система распределения ресурсов должна включать модуль мониторинга, осуществляющий сбор метрик загрузки процессора, памяти и сетевых ресурсов; модуль прогнозирования, анализирующий исторические данные; модуль принятия решений, выбирающий оптимальную стратегию распределения; а также механизм обратной связи, корректирующий политику управления. Такой замкнутый контур управления обеспечивает адаптивность и устойчивость облачной инфраструктуры.

Перспективным направлением является разработка гибридных алгоритмов, сочетающих классические методы оптимизации и машинное обучение.

Дополнительный интерес представляет интеграция объяснимого искусственного интеллекта, позволяющего анализировать и интерпретировать принятые системой решения. В условиях роста масштабов облачных систем также актуальны исследования в области энергоэффективного управления и распределённых моделей обучения.

Таким образом, разработка интеллектуальных методов и алгоритмов адаптивного распределения вычислительных ресурсов является важной задачей современной информатики. Применение комплексных и самообучающихся механизмов управления позволяет повысить эффективность использования инфраструктуры, снизить эксплуатационные расходы и обеспечить высокое качество обслуживания пользователей в условиях постоянно изменяющейся среды.

Список литературы:

1. Buyya R., Broberg J., Goscinski A. Cloud computing: principles and paradigms. — Hoboken: John Wiley & Sons, 2011. — 664 p.
2. Erl T., Puttini R., Mahmood Z. Cloud computing: concepts, technology & architecture. — Boston: Prentice Hall, 2013. — 528 p.
3. Mao M., Humphrey M. A performance study on the VM startup time in the cloud // Proceedings of the IEEE International Conference on Cloud Computing. — 2012. — P. 423–430.