

В.Д. Кисляк, 3 курс**Научный руководитель – Я.В. Богатко, ассистент****Полесский государственный университет**

В статье рассматриваются основные методы балансировки нагрузки в распределенных системах, их преимущества и недостатки. Анализируются реализации этих методов в белорусских и российских IT-решениях, а также их математическое обоснование.

С ростом объемов данных и количества пользователей распределенные системы становятся основой современных IT-инфраструктур. В Беларуси это особенно актуально в связи с развитием облачных сервисов (Яндекс.Облако, beCloud [3]) и государственных информационных систем [1, с.8]. Одна из ключевых задач - балансировка нагрузки, обеспечивающая равномерное распределение запросов между серверами.

Основные методы балансировки нагрузки включают в себя статистические и динамические алгоритмы.

Статические алгоритмы:

Round Robin (циклический перебор) - запросы распределяются по очереди. Используется, например, в белорусской системе "АИС Реестр населения" для обработки обращений граждан [5, с.78];

Weighted Round Robin (с весовыми коэффициентами) - мощные серверы получают больше запросов. Применяется в российской системе "Сбербанк-Онлайн" [4, с.56].

Динамические алгоритмы:

Least Connections (наименьших соединений) - сервер с минимальной нагрузкой получает новый запрос. Используется в белорусской платформе "МАРТ" для электронных торговых площадок [5, с.82];

IP Hash (хэширование по IP) - запросы от одного пользователя всегда направляются на один сервер. Применяется в российских SEO-сервисах (например, "Метрином") [4, с.61].

Для анализа эффективности используется модель М/М/п (математическая модель системы массового обслуживания, широко используемая для анализа производительности распределённых систем, серверов и сетевых инфраструктур) [2, с.125]:

λ - интенсивность потока запросов (например, 1000 запросов/с в системе "ГАС "Правосудие" РБ [1, с.15]),

μ - скорость обработки (например, 200 запросов/с на сервер),

n - количество серверов.

Условие стабильности системы:

$$\rho = \lambda / \mu n < 1$$

При $\rho \geq 1$ возникают очереди, что приводит к сбоям.

Примеры реализации в Беларуси и России:

- beCloud (Беларусь) - использует Weighted Least Connections для облачных ВМ [3].

- Яндекс.Балансер (Россия) - комбинирует IP Hash и Response Time для видеосервиса "Кинопоэзия" [4, с.72].

- Белорусская АИС "Электронное обращение" - Round Robin для равномерного распределения заявок по регионам [5, с.85].

Для простых сервисов (новости, справки) достаточно Round Robin [5, с.88].

Для сложных систем (банки, стриминг) лучше использовать динамические методы (Least Connections) [4, с.75].

Белорусские разработки (например, "БелХард") часто сочетают несколько алгоритмов для повышения отказоустойчивости [1, с.20].

Список использованных источников

1. Государственный стандарт Республики Беларусь СТБ 34.101.47-2023 "Требования к отказоустойчивости информационных систем". – Введ. 2023-01-01. – Минск: Госстандарт, 2022. – 45 с.

2. Клейнрок, Л. Теория массового обслуживания / Л. Клейнрок; пер. с англ. – Москва: Машиностроение, 1979. – 432 с.

3. Официальная документация платформы beCloud [Электронный ресурс]. – Режим доступа: <https://becloud.by/docs> – Дата доступа: 10.05.2024.

4. Петраченко, А.В. Облачные технологии в России: опыт Яндекса / А.В. Петраченко. – Москва: Изд-во МГТУ, 2022. – 180 с.

5. Савицкий, К.Л. Распределенные системы: теория и практика / К.Л. Савицкий. – Минск: БГУ, 2021. – 245 с.