

УДК: 621.126

ОПТИМІЗАЦІЯ ПРОПУСКНОЇ ЗДАТНОСТІ В СИСТЕМАХ ПАКЕТНОЇ ОБРОБКИ ДАНИХ

Денисов Р.К.

Науковий керівник – канд. техн. наук, доц. Єсілевський В.С.
Харківський національний університет радіоелектроніки, каф. ПМ,
м. Харків, Україна
e-mail: roman.denysov@nure.ua

This paper proposed an optimization of throughput in the Batch-Processing Systems, specifically in a closed queueing system with n clients and m servers. Generally, clients are producing jobs that are processed in batches by parallel servers. The main characteristics that impact system throughput are batch size and corresponding speedup function. In such systems, the speedup function has the sub-additive property that means that infinitely increased computational resources will not affect the efficiency linearly. In fact, the problem of throughput optimization leads to searching for the optimal batch size parameter. The process of finding the batch size parameter could take up to multiple weeks in existing big tech systems. This paper has strived to find optimal throughput in $O(1)$ time by referring to the mean-field model (MFM).

Коли в комерційну систему починає надходити велика кількість запитів, з'являється ряд проблем: система не витримує кількість запитів, через що падає і втрачаються дорогоцінні дані, даних може приходити набагато більше, ніж здатна обробити звичайна система, або дані можуть оброблятися надзвичайно довго. Ключовою технікою оптимізації у подібних високонавантажених системах обробки даних є пакетна обробка (batch processing), що полягає у формуванні вхідних даних у пакети із послідувочою їх обробкою як одного об'єкту [1].

Batch size (BS) – це параметр системи пакетної обробки, який регулює кількість задач, які можуть оброблюватися одночасно [2]. На практиці, для пошуку оптимального значення BS_0 використовують чисельні методи, методи оптимізації або симуляція моделі системи. Проблема в тому, що під час великого об'єму даних ці методи втрачають свою ефективність й пошук цього параметру може займати великий обсяг часу. Якщо параметр $BS \ll BS_0$, то система буде обробляти набагато менше даних, ніж до нас надходить, і при такому розкладі послідовність вхідних даних $\{X_n\} \rightarrow \infty$, це призведе до падіння нашої системи. У випадку, коли $BS \gg BS_0$, час на обробку одного пакету $t_{\text{пак}} \rightarrow \infty$, й це знову призведе до падіння нашої системи.

Метою даної роботи є пошук асимптотично оптимальної пропускної здатності для закритої системи за $O(1)$ часу. Це дозволить побудувати систему, яка буде взмозі обробляти великий потік даних за скінченну кількість часу без їх втрати, що є актуальною проблемою для сучасних систем.

У ході дослідження було встановлено, що модель середнього поля (mean-field model) має унікальну, глобальну стаціонарну точку, яку можна знайти, і яка характеризує асимптотичну пропускну здатність системи, як функцію від BS [3]. Ідея даної моделі полягає у тому, що поведінка кожного елемента системи може бути описана середнім значенням його характеристик. Вона дозволяє замінити складний мікрорівень взаємодії на більш простий макро-рівень, що спрощує аналіз системи. До цього стандартом галузі було використання Марковського підходу, де час пошуку BS оцінювався як $O(n^4)$. Влаштувавши для тестового полігону чисельні випробування, було встановлено асимптотичну оптимальну пропускну здатність в $O(1)$ часу для нової моделі.

Дана робота має практичне застосування в сучасних високонавантажених системах обробки даних, де важливо забезпечити оптимальну пропускну здатність і ефективність обробки завдань. Шляхом моделювання системи як закритої мережі черг, де партії спостерігають субаддитивне прискорення виконання, отримано оптимальну пропускну здатність як функцію розміру партії для n клієнтів та m серверів. Розглянута система схожа на стандартні системи баз даних, де клієнти чекають на результат вхідного запиту для генерації наступного. Використання моделі середнього поля дозволяє точно прогнозувати ефективність системи в асимптотичних умовах та розробляти оптимальні стратегії управління ресурсами. Загалом, ця робота відкриває нові перспективи для оптимізації обробки даних у великих та складних системах, що має велике значення для подальшого розвитку сучасних технологій та підвищення їхньої продуктивності.

Список використаних джерел:

1. <https://blog.k2datascience.com/batch-processing-apache-spark-a67016008167>.
2. <https://www.sciencedirect.com/topics/mathematics/mean-field-model>.
3. Forest Baskett, K Mani Chandy, Richard R Muntz, Fernando G Palacios, et al. 1975. Open, closed, and mixed networks of queues with different classes of customers. *J. ACM* 22, 2 (1975), 248–260.
4. Yesilevskyi, V. et al. (2022), "Development of a video processing module for the task of air object recognition based on their contours", *Innovative Technologies and Scientific Solutions for Industries*, No. 3 (21), P. 16–25. DOI: <https://doi.org/10.30837/ITSSI.2022.21.016>
5. K Mani Chandy and Alain J Martin. 1983. A characterization of product-form queuing networks. *J. ACM* 30, 2 (1983), 286–299.
6. X. Wen, B. Yang, Y. Chen, L. E. Li, K. Bu, P. Zheng, Y. Yang, and C. Hu. 2016. RuleTris: Minimizing Rule Update Latency for TCAM-Based SDN Switches. In *Proc. IEEE Int. Conf. Dist. Comput. Sys.* 179–188.