

Шевченко Є. – гр. ІІІ-51м
НТУУ "КПІ" ім. І. Сікорського

БАЛАНСУВАННЯ НАВАНТАЖЕННЯ МІЖ КЛАСТЕРОМ СЕРВЕРІВ З УРАХУВАННЯМ ТИПУ ЗАДАЧ

Науковий керівник: к.т.н., ст. викл. Ліщук К.І.

Shevchenko I.
NTUU "Igor Sikorsky Kyiv Polytechnic Institute"

METHOD OF LOAD BALANCING BETWEEN SERVERS BASED ON THE TYPE OF TASKS

Supervisor: Lischuk K.

Ключові слова: балансування, навантаження.
Keywords:balancing, load.

З кожним роком кількість нових інтернет-ресурсів, які виконують різноманітні задачі зростає експоненціально відносно кількості нових користувачів в мережі Інтернет. Лише за 2016 рік було зафіксовано приріст користувачів мережі близько семи відсотків. Але зростає не тільки кількість нових ресурсів а й навантаження на вже існуючі. Великі за обсягом трафіку компанії для підвищення надійності та доступності своїх ресурсів використовують так звані методи горизонтального масштабування, а саме збільшення кількості серверів для обробки даних.

Для розподілення навантаження між великими кластерами серверів використовується проміжне програмне забезпечення, так звані балансувальники (LoadBalancers). В переважній більшості випадків такі балансувальники працюють методом простого й почергового перебору всіх доступних серверів, тобто при надходженні нової задачі балансувальник нічого не враховує окрім черговості серверів.

Опис проблеми. Уявімо ситуацію коли у розпорядженні простого балансувальника є 5 серверів які займаються обробкою ресурсоемких задач, задачі по рівню їх складності можуть займати від кількох секунд до кількох годин роботи серверу в режимі максимального завантаження. Метод простого перебору серверів буде не оптимальним в разі якщо перша задача була максимально ресурсоемна а всі наступні були прості та швидкі по часу виконання. В такому випадку шоста задача буде довгий час очікувати вирішення першої, хоча задачі з другої по п'яту до цього моменту були вже вирішені а сервери готові отримувати наступні задачі.

Відомі рішення. Одним з найпопулярніших рішень цієї проблеми є введення максимального часу відповіді серверу про прийняття задачі на вирішення (timeout). Тобто це час після якого балансувальник спробує передати задачу наступному серверу. Такий підхід є тільки частковим вирішенням проблеми тому що: 1. Вводиться певний timeout, під час якого задача буде простоювати, а отже час отримання відповіді збільшиться на невідому величину. 2. Додаткове навантаження на оперативну пам'ять балансувальника.

Пропоноване рішення. Для більш ефективного розподілу навантаження між серверами пропонується розробити таке програмне забезпечення балансувальника яке не буде зберігати в пам'яті нічого що стосується самої задачі а лише буде підтримувати певний протокол правил за якими дана задача буде розподілена по кластеру. Даний підхід можливий тільки для балансування на прикладному рівні, в свою чергу це означає що програмне забезпечення буде окремо налаштовуватися для кожного набору задач.