

Streszczenie notatki "Analiza wydolności serwisu" oraz podrozdziału "Skalowanie geograficzne"

Julia Majkowska

Analiza wydolności serwisu

Będziemy traktować serwis jako system z kolejką o nieskończonej pojemności, w której trzymane są przychodzące zapytania. Jak proces jest w stanie przetworzyć pierwsze w kolejce zapytanie to przyjmuje takie zapytanie, wykonuje je i wysyła odpowiedź.

Załóżmy, że nasz system przyjmuje λ zapytań na sekundę i jest w stanie przetworzyć μ zapytań na sekundę. Najbardziej istotną cechą tego systemu jest R , czyli jak długi jest czas od przyjscia zapytania do wysłania odpowiedzi. Policzmy najpierw przez jaką część czasu w całym systemie (w kolejce i w trakcie przetwarzania) jest dokładnie k zapytań.

$$p_k = (1 - \frac{\lambda}{\mu})(\frac{\lambda}{\mu})^k$$

Zdefiniujmy sobie utylizację jako :

$$U = \sum_{k \geq 0} p_k = 1 - p_0 = \frac{\lambda}{\mu}$$

Korzystając z prawa Little'a można wyprowadzić, że

$$R = \frac{\text{średnia liczba zapytań w systemie}}{\text{średnia przepustowość systemu}} = \frac{\sum_{k \geq 0} k p_k}{U\mu + 0(1 - U)} = \frac{1}{\mu(1 - U)},$$

gdzie ostatnie przejście można wykazać przy pomocy prostych przekształceń. Z tego wynika, że gdy utylizacja jest bliska jedynki czas oczekiwania może być bardzo duży. W takiej sytuacji konieczne jest skrócenie czasu przetwarzania zapytania.

Skalowanie geograficzne

Wiele systemów rozproszonych zostało zaprojektowanych dla lokalnych sieci opartych na komunikacji synchronicznej, gdzie klient wstrzymuje działanie do czasu otrzymania odpowiedzi na zapytanie. Przy komunikacji na większą odległość opóźnienia przy przesyłaniu informacji są znacznie większe, dlatego zbudowanie systemu opartego na komunikacji synchronicznej na dużą odległość wymaga zastosowania innych rozwiązań. Połączenie na dużą odległość jest mniej

stabilne od lokalnego i ma bardziej ograniczoną przepustowość łącza. Dodatkowo jeśli komponenty są oddalone od siebie, aby koordynować wielopunktową komunikację, należy stworzyć osobne (również skalowalne) systemy nazewnicze i katalogowe.

Na podstawie "Distributed Systems. Third edition", Maarten van Steen, Andrew S. Tanenbaum, s 16 -18