

Streszczenie
wydzielonego fragmentu „Being scalable”
z „Distributed Systems 3rd edition 2017”,
strony 15-16,
Maarten van Steen, Andrew S. Tanenbaum

streszczenie wykonał
Piotr Łężny

Bycie skalowalnym

Coraz większa powszechność dostępu do Sieci sprawia, że odchodzimy od dużych komputerów stacjonarnych, dostosowanych do korzystania z określonych aplikacji, na rzecz mniejszych urządzeń, jak tablety i smartfony, na których możemy korzystać z „chmurowych” odpowiedników tych programów. Skalowalność jest bardzo ważna dla twórców systemów rozproszonych.

Wymiary skalowalności

Skalowalność systemu może być mierzona w co najmniej trzech różnych wymiarach:

Skalowanie rozmiarowe: System może być skalowany ze względu na jego rozmiar. Oznacza to, że możemy dodać więcej użytkowników i zasobów do systemu bez zauważalnej zmiany w wydajności.

Skalowanie geograficzne: Geograficznie skalowalny system to taki, w którym użytkownicy i zasoby mogą leżeć daleko od siebie, ale duże opóźnienia w komunikacji nie są często zauważane.

Skalowanie administracyjne: Administracyjnie skalowalny system to taki, którym można łatwo zarządzać nawet jeśli obejmuje wiele niezależnych, administracyjnych organizacji.

Skalowanie rozmiarowe

Usługi scentralizowane, czyli korzystające z pojedynczego serwera w systemie rozproszonym, często spotykają się z problemami przy większej liczbie użytkowników. Główne trzy przyczyny spadku wydajności w takiej sytuacji to:

- moc obliczeniowa, ograniczona mocą procesorów
- pojemność oraz przepustowość pamięci, włączając w to operacje wejścia/wyjścia
- ograniczona przepustowość połączenia sieciowego między użytkownikiem, a serwerem

Rozważmy najpierw moc obliczeniową. W przypadku usług wykonujących wymagające i skomplikowane obliczenia, łatwo można zauważyć, że obsługując wiele żądań jednocześnie, pojedyncze urządzenie spotka się z problemami związanymi z wydajnością.

W przypadku operacji wejścia/wyjścia doświadczymy podobnych problemów, ale z innych powodów. Typowym przykładem jest źle zaprojektowana wyszukiwarka. Problem polega na tym, że musimy dopasować zapytanie do całego zestawu danych. Przetwarzanie sporej ilości danych będzie problemem nawet przy użyciu zaawansowanego indeksowania. Dodanie większej liczby szybkich dysków może nie wystarczyć przy użyciu pojedynczego serwera.

Połączenie sieciowe, może być problemem na przykład w usługach do przesyłania strumieniowego filmów w wysokiej jakości. Bardzo szerokie pasmo będzie niewystarczające przy bardzo dużej liczbie użytkowników.