

MACIEJ ROSZKOWSKI

Zachodniopomorski Uniwersytet Technologiczny¹

SKALOWALNOŚĆ SYSTEMÓW INFORMATYCZNYCH W OBSZARZE E-ADMINISTRACJI

Streszczenie

Artykuł prezentuje skalowalność systemów informatycznych w kontekście systemu usług e-administracji. Zaprezentowano opis systemów informatycznych, identyfikację wizualną profili skalowalności systemu informatycznego oraz dwa prawa skalowalności (prawo skalowalności Amdahla i prawo skalowalności uniwersalnej). Przedstawiono również przykład skalowalności aplikacji w oparciu o przedstawione modele skalowalności.

Słowa kluczowe: prawo skalowalności Amdahla, prawo skalowalności uniwersalnej.

Wprowadzenie

Elektroniczna administracja wykorzystuje technologie informacyjne i telekomunikacyjne w administracji publicznej. E-administracja ma za zadanie zachęcić obywateli do większego zaangażowania w procesy demokratyczne poprzez poprawienie jakości świadczonych usług publicznych, a jednocześnie wspierać funkcjonowanie mechanizmów państwa. Elektroniczna administracja niesie za sobą zmiany i usprawnienia organizacyjne oraz optymalizację procesów administracyjnych pod kątem efektywności.

Według badań Capgemini z 2012 roku (na zlecenie Komisji Europejskiej) mieszkańcy Unii Europejskiej deklarują, że korzystając z e-administracji najczęściej wykorzystują następujące usługi (Capgemini 2014, s. 21): pośrednictwo w poszukiwaniu pracy, rozliczenie podatku dochodowego, złożenie dokumentów

¹ Wydział Informatyki, Katedra Inżynierii Systemów Informatycznych.

o przyjęcie na studia, świadczenia emerytalne, zmiana miejsca pobytu, dostęp do zasobów bibliotek publicznych i wiele innych. Wszystkie te usługi e-administracji są dostępne online (przez Internet) przez całą dobę, we wszystkie dni tygodnia.

Użytkownicy e-administracji wykorzystują możliwość dostępu do usług w dogodnym dla siebie momencie. Często jednak, pod koniec dat granicznych (ustawowych rozliczeń z administracją rządową lub samorządową), zdarzają się okresy nasilonego wykorzystania usług e-administracji. Bardzo duża liczba połączeń użytkowników, którzy na ostatnią chwilę pragną skorzystać z przysługującego im prawa realizacji usługi e-administracji, powoduje, że są często rozczarowani: powolnym działaniem systemu, przerwami uniemożliwiającymi dokończenie rozpoczętej usługi, czasową niedostępnością usług lub wręcz niedostępnością platformy informatycznej. Systemy operacyjne, na których rezydują usługi e-administracji, nie są w stanie dostosować się do nasilenia się usług i odmawiają posłuszeństwa.

Przykładem awarii związanej z brakiem skalowalności może być funkcjonowanie platformy ePUAP podczas składania deklaracji do ZUS o chęci dalszego oszczędzania na emeryturę w OFE (do 31 lipca 2014). System ePUAP umożliwia założenie profilu PUE na portalu ZUS. W połowie okresu składania deklaracji (po 2 miesiącach) złożono ponad 200 tysięcy deklaracji, a pod koniec okresu (w ostatnim tygodniu) dziennie składano około 100 tysięcy deklaracji. System logowania do systemu ePUAP był w tym czasie wielokrotnie niedostępny (Epuap 2014), m.in. ze względu na zwiększone obciążenie, skutecznie uniemożliwiając realizację usługi bez dodatkowego potwierdzenia profilu PUE w placówce ZUS. Ostatecznie w OFE pozostało ponad 2,5 mln osób (z 14 mln dotychczasowych członków funduszy). Większość osób deklarację przesłała listownie w ostatnim tygodniu (około 1 mln osób).

Celem niniejszej publikacji jest przedstawienie idei skalowalności systemu informatycznego w kontekście systemu usług e-administracji. Na początku zostanie zaprezentowane pojęcie skalowalności i jego zastosowanie w systemach informatycznych, w których wzrasta obciążenie. W kolejnej części artykułu zostanie przedstawiona identyfikacja wizualna profili skalowalności systemu informatycznego. Zostaną również przedstawione dwa prawa skalowalności: prawo skalowalności Amdahla i prawo skalowalności uniwersalnej, które umożliwiają wyrażenie skalowalności systemu w postaci modelu matematycznego. Na końcu zostanie zaprezentowany przykład skalowalności aplikacji oraz interpretacja modelu skalowalności Amdahla i modelu skalowalności uniwersalnej.

1. Skalowalność

Skalowalność (ang. *scalability*) to zdolność systemu do dostosowywania się do rosnącego nasilenia się usług (Silberschatz 2005, s. 664). System skalowalny dostos-

sowuje się do wzrastającego obciążenia w sposób bardziej harmonijny niż ma to miejsce w przypadku systemu nieskalowanego. Wraz ze wzrostem obciążenia wydajność maleje, aż do osiągnięcia stanu nasycenia. System, który jest skalowany, powinien umożliwiać funkcjonowanie w wysokim obciążeniu usługami oraz powinien w łatwy sposób umożliwiać rozrastanie się całego systemu w sposób harmonijny.

Skalowalność umożliwia sprawdzenie, w jaki sposób zakładana wydajność będzie skalowalna, jeżeli zmianie ulegną obciążenia lub zasoby (np. rdzenie procesora, procesy, wątki).

W przypadku badania istniejącego systemu znajdującego się pod obciążeniem analizę wydajności należy poprzedzić charakterystyką obciążenia i pomiarem osiąganego wydajności. Natomiast w celu przetestowania obciążenia większego niż stosowane w produkcji – analizę wydajności poprzedza testowanie symulowanego obciążenia. W obydwu przypadkach modelowanie analityczne pozwala przewidzieć wydajność w oparciu o wyniki pomiaru lub symulację obciążenia.

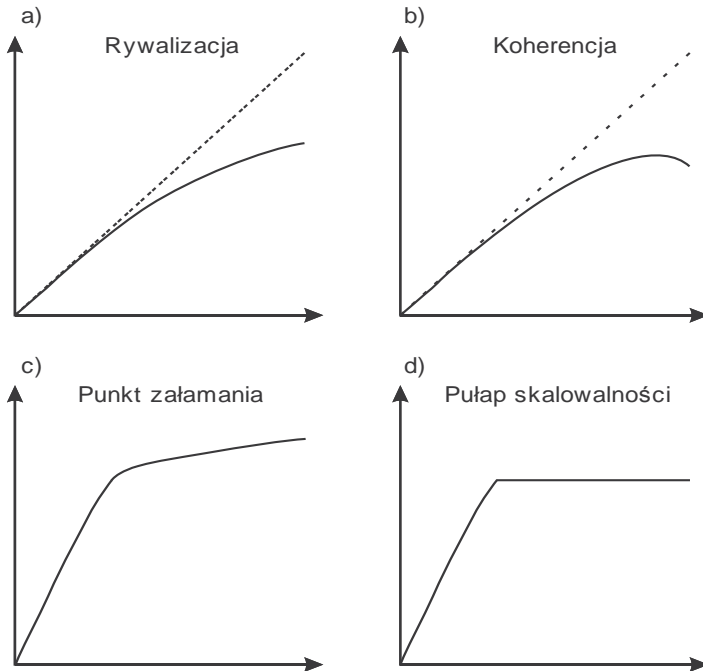
2. Identyfikacja wizualna

W oparciu o przeprowadzone eksperymenty można zbudować wykresy pokazujące wydajność systemu. Większość profili skalowalności można zidentyfikować wizualnie, bez wnikania w model matematyczny (rys. 1).

Identyfikacja wizualna wydajności systemu jest możliwa za pomocą prostych dwuwymiarowych wykresów. Oś X przedstawia kierunek skalowania, natomiast oś Y – osiągniętą wydajność (rys. 1).

Rodzaje profili skalowalności (Gregg 2014, s. 59):

- Profil skalowalności liniowej przedstawia wzrost wydajności wprost proporcjonalny do skalowanego zasobu. Skalowalność liniowa przedstawia funkcja $y = x$ (jest oznaczoną linią przerywaną na rys. 1a i 1b).
- Profil rywalizacji (rys. 1a) pokazuje spadek efektywności skalowania w wyniku rywalizacji między współdzielonymi zasobami. Jest to spowodowane faktem, że niektóre zasoby mogą być używane szeregowo.
- Profil koherencji (rys. 1b) pokazuje spadek efektywności skalowania, przechodzący w spadek wydajności. Koszt, jaki poniesiemy w przypadku stałego wzrostu wydajności, może być o wiele wyższy niż zalety, jakie niesie za sobą skalowanie.
- Profil punktu załamania (rys. 1c) pokazuje, że można wyznaczyć punkt załamania, w którym zmienia się wzorzec skalowalności.
- Profil pułapu skalowalności (rys. 1d) pokazuje, że osiągnięto pewne ograniczenie i dalsze skalowanie nie przyniesie wzrostu wydajności. Najczęściej ma to związek z wykryciem „wąskiego gardła” lub ograniczenia w badanym systemie.



Rys. 1. Profile skalowalności

Źródło: opracowanie własne.

3. Prawo skalowalności Amdahla

Prawo skalowalności Amdahla jest najczęściej omawiane w kontekście przeprowadzania obliczeń równoległych, w celu obliczenia maksymalnego wzrostu szybkości obliczeń przy użyciu wielu procesorów jednocześnie. Prawo to pozwala opisać skalowalność systemu, biorąc pod uwagę szeregowe dane obciążające system, które nie skalują się równoległe (rys. 2b). Może zostać zastosowane do analizy skalowalności procesorów, wątków, obciążeń i innych zasobów. Prawo skalowalności Amdahla opisuje wzór:

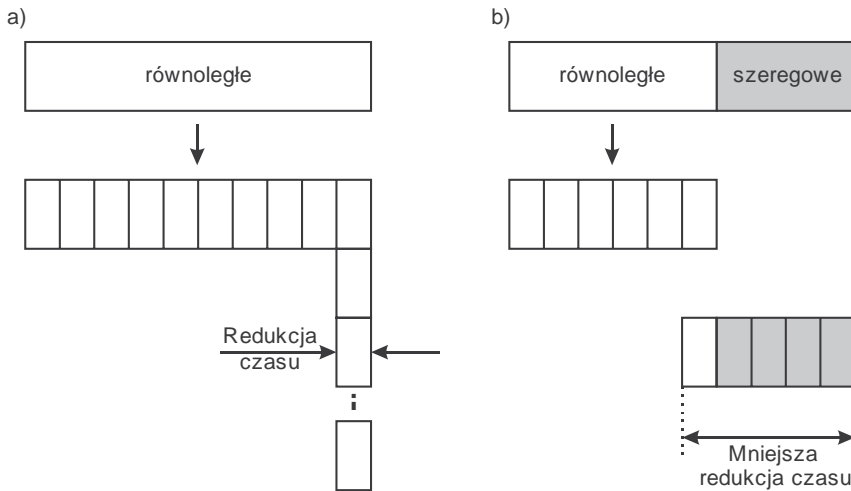
$$C(N) = \frac{N}{1 + \alpha(N - 1)}$$

gdzie:

C(N) – względna pojemność,

N – wymiar skalowania (np. liczba procesorów, wątków, obciążenie),

 α – parametr określający stopień szeregowości oraz odchylenie od skalowalności liniowej, $\alpha \in (0; 1)$.



Rys. 2. Redukcja czasu w dwóch wariantach: idealna równoległość, dane przetwarzane równoległe i szeregowe (prawo Amdahla)

Źródło: opracowanie własne.

Prawo Amdahla pokazuje, że niemożliwe jest osiągnięcie idealnej równoległości (rys. 2a), ponieważ istnieją pewne części, które mogą zostać przetworzone jedynie szeregowo (szara część na rys. 2b).

Aby zastosować prawo skalowalności Amdahla, należy zebrać dane dla wymiaru skalowania N za pomocą mikrotestów wydajności (eksperymentalnie w przypadku prostego obciążenia lub za pomocą generatorów obciążenia). W celu ustalenia parametru α należy przeprowadzić analizę regresji.

4. Prawo skalowalności uniwersalnej

Prawo skalowalności uniwersalnej (ang. *Universal Scalability Law*) nawiązuje do prawa skalowalności Amdahla, uwzględniając jednocześnie opóźnienie koherencji β (Gunther 2007, s. 56). Koherencja została pokazana jako profil skalowalności (rys. 1b). Prawo skalowalności Amdahla jest szczególnym przypadkiem prawa skalowalności uniwersalnej dla $\beta=0$. Prawo skalowalności uniwersalnej opisuje wzór:

$$C(N) = \frac{N}{[1 + \alpha(N - 1)] + \beta N(N - 1)}$$

gdzie:

$C(N)$, N , α – jak w prawie skalowalności Amdahla,

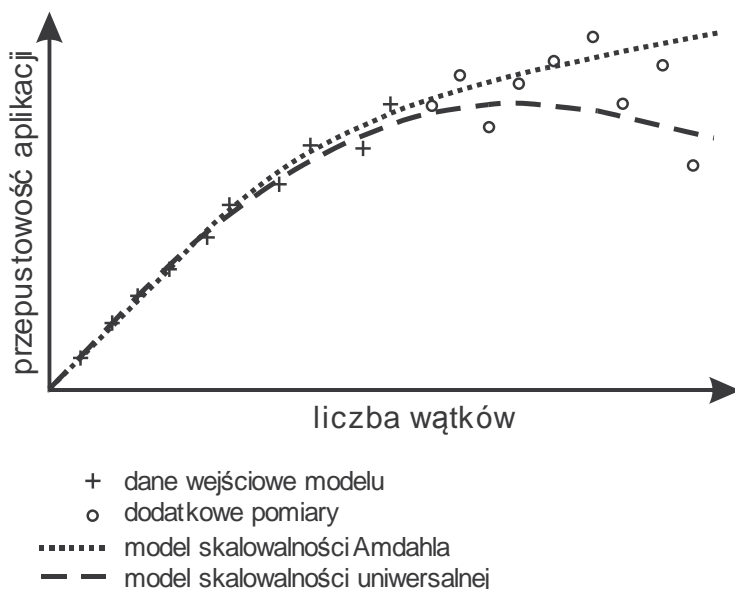
β – parametr koherencji (opóźnienie dla danych, aby stały się spójne).

5. Przykład skalowalności aplikacji

Dla pewnej aplikacji postanowiono zmierzyć przepustowość (ilość wykonywanej pracy, szybkość przekazywanych danych) w stosunku do liczby skalowanych wątków. Dane wejściowe dla obydwu modeli (modelu skalowalności Amdahla i modelu skalowalności uniwersalnej) zostały zaznaczone krzyżykami, dalsze pomiary dodatkowe systemu – kółkami, model skalowalności Amdahla – krzywą wykropkowaną, model skalowalności uniwersalnej – krzywą przerywaną (rys. 3).

Model skalowalności Amdahla dla aplikacji pokazuje, że dodawanie kolejnych wątków nie powoduje już zwiększenia wydajności na oczekiwanym poziomie. Może być to dla nas sygnałem, że należy aplikację przebadać i skonfigurować system w poszukiwanej optymalnej liczbie wątków.

Model skalowalności uniwersalnej dla aplikacji pokazuje, że w momencie kiedy poziom wykorzystania procesora osiągnie 100% (w wyniku zwiększającej się ciągle liczby wątków), wydajność zacznie spadać. Jest to spowodowane kolejkowaniem zadań i wzrastającym opóźnieniem. W momencie osiągnięcia maksymalnej wydajności przepustowość będzie malała (wraz ze zwiększającą się liczbą kolejnych wątków). Większa liczba wątków do obsłużenia spowoduje większą liczbę operacji przełączania kontekstu, a więc większe zużycie zasobów procesora i mniejszą liczbę wykonanych na bieżąco zadań.



Rys. 3. Modele skalowalności

Źródło: opracowanie własne.

Podsumowanie

Skalowalność jest bardzo dobrym mechanizmem umożliwiającym sprawdzenie systemów informatycznych pod kątem wydajności przy wzrastającym obciążeniu. Problem obciążenia w dużym stopniu dotyczy usług e-administracji, gdzie są wyznaczone graniczne daty rozliczeń, przed którymi wzrasta natężenie wykorzystania tych usług. Możemy obliczyć, jak duże muszą być zasoby systemu informatycznego, aby był w stanie obsłużyć rosnące nasilenie się usług. Za pomocą profili skalowalności możemy wyrazić wydajność systemu, bez zagłębiania się w model matematyczny. Identyfikacja wizualna profili skalowalności może być łatwa i efektywna, ale nie zastąpi modelu matematycznego. Bardzo często model systemu może odbiegać od danych, co może oznaczać problem w samym modelu, błędne zrozumienie systemu lub problem w rzeczywistej skalowalności systemu.

Literatura

1. Gregg B. (2014), *Systems Performance: Enterprise and the Cloud*, Prentice Hall, Michigan.
2. Gunther N.J. (2007), *Guerrilla Capacity Planning*, Springer, Heidelberg.
3. Silberschatz A., Galvin P.B., Gagne G. (2005), *Podstawy systemów operacyjnych*, WNT, Warszawa.
4. www.capgemini.com/resource-file-access/resource/pdf/egov_benchmark_2012_insight_report_final.pdf (2014), eGovernment Benchmark 2012 – INSIGHT report.
5. www.epuap.gov.pl/wps/portal/E2_Aktualnosci (2014).

A SCALABILITY OF IT SYSTEMS IN THE AREA OF THE E-GOVERNMENT

Summary

This article presents a scalability of IT systems in the area of the e-government. The paper described IT systems, visual identification of IT systems' scalability profiles and two scalability laws (Amdahl's Law of Scalability and Universal Law Scalability). It also shows example of scalability of application based on the presented scalability models.

Keywords: Amdahl's Law of Scalability, Universal Law Scalability.

Translated by Maciej Roszkowski

