Porównanie wydajności chmur i superkomputerów

1. **Model rzeczywistości**

Dany jest system obsługi zadań. System ten pracuje w trybie online, obsługując strumień zadań przedkładanych przez użytkowników tego systemu. Zadania mają określony czas przedłożenia i rozmiar, zdefiniowany w jednostkach czasu niezbędnych do wykonania każdego z zadań przez system. Częstość przedkładania zadań i rozmiar zadań są wartościami losowymi.

Zmienność odstępów czasu między przedkładaniem kolejnych zadań ma być modelowana za pomocą fazowych rozkładów bimodalnych reprezentujących okresy dużych i małych obciążeń systemu. Rozkład bimodalny ma się składać z dwóch rozkładów wykładniczych różniących się wartościami średnimi. Fazowość ma symulować czasową rozdzielność okresów dużego (rozkład wykładniczy o małej średniej wielkości odstępów czasu) i małego obciążenia systemu (rozkład wykładniczy o dużej średniej wielkości odstępów czasu). Zmienność wielkości zadań ma być modelowana rozkładu bimodalnego. Rozkład bimodalny ma się składać z dwóch rozkładów Erlanga różniących się wartościami średnimi: rozkład z dużą wartością średnią dla dużych zadań i rozkład z małą wartością średnią dla małych zadań.

System obsługi zadań jest chmurą o danej liczbie węzłów N. Liczba węzłów jest zmiennym parametrem modelu. Eksperymenty mają być przeprowadzone dla wybranych wartości N z przedziału od 1 do 100. Eksperymenty maja przeprowadzone dla kilku protokołów obsługi zadań: JNQ (Join Null Queue) dla alokacji i FCFS dla szeregowania oraz JSQ (Join Shortest Queue) dla alokacji i PS dla szeregowania.

Wynikiem badań mają być wykresy średnich czasów odpowiedzi, ale również ich składowych średnich czasów przetwarzania i opóźnienia, w funkcji obciążenia systemu, współczynnika zmienności (odchylenie standardowe/wartość średnią) czasów między przedkładaniem zadań, współczynnika zmienności rozmiarów zadań **oraz** sposobu porządkowania w strumieniu wejściowym czasów między przedkładaniem zadań za pomocą generowania dwufazowych cykli zadań..

Dwie fazy cyklu mają różnić się poziomem „jednorodności”. Poziom jednorodności ma być niezależnym zmiennym parametrem eksperymentów. Dla każdej z dwóch faz czasy między zadaniami są generowane z losowo wybranego rozkładu składowego rozkładu bimodalnego, czyli o mniejszej albo większej średniej wartości czasu. Dla całkowicie jednorodnych faz, czasy między zadaniami w pierwszej fazie są generowane z rozkładu o mniejszej średniej ze 100% pewnością, a z rozkładu o większej średniej z prawdopodobieństwem równym zero. Analogicznie czasy między zadaniami w drugiej fazie są generowane z rozkładu o mniejszej średniej z prawdopodobieństwem równym zero, a z rozkładu większej o średniej z prawdopodobieństwem równym jeden. Dla faz niejednorodnych, powyższe prawdopodobieństwa należą do przedziału <0.5; 1). Na przykład, w pierwszej fazie wybór pierwszego rozkładu 90%, a drugiego 10%, i wtedy w fazie drugiej wybór pierwszego rozkładu 10%, a drugiego 90%. Przypadek dla którego wartości wszystkich prawdopodobieństw są równe 0.5 odpowiada sytuacji braku faz.

Należy przeprowadzić eksperymenty dla kilku poziomów jednorodności grup z przedziału <0.5; 1>, np. dla 0.5; 0.6; 0.7; 0.8; 0.9 i 1 (przedział <0; 0.5) nie i jest przedmiotem naszego zainteresowania). Wymienione wartości odpowiadają prawdopodobieństwu generowania czasu z rozkładu o mniejszej średniej dla pierwszej fazy i jednocześnie prawdopodobieństwu generowania czasu z rozkładu o większej średniej dla drugiej fazy. Prawdopodobieństwa generowania czasu z rozkładu o większej średniej dla pierwszej fazy i mniejszej średniej dla drugiej są wyznaczane jako uzupełnienie do jedynki.

Dla zmienianych wartości jednego z parametrów np. obciążenia, pozostałe parametry mają być stałe.

**Jak zachować średnią wartość przy zmianach odchylenia standardowego czasów przedkładania zadań?** Czyli np. stałe obciążenie, przy zmianach współczynnika zmienności czasów przedkładania zadań.

**Rozwiązanie nr 1** – Symetrycznie rozsuwać/zsuwać wartości średnie rozkładów składowych rozkładu bimodalnego względem wspólnej średniej wartości czasu, tak żeby po rozsunięciu zachować tę wartość średnią:

μ = %1\*μ1 + %2\*μ2 = %1’\*μ1’ + %2’\*2μ2’

gdzie: wartości %1 i %2 odpowiadają względnej liczebności wartości należących do rozkładów składowych – rozkładu bimodalnego. Dla równolicznych podzbiorów %1 = %2 = 50%:

μ = 0,5μ1 + 0,5μ2 = 0,5μ1’ + 0,5μ2’ = ((μ1 + Δ) + (μ2 - Δ))/2 = (1/λ’1 + 1/λ’2)/2 = ((1/λ1 + Δ) + (1/λ2 - Δ)/2)

gdzie: Δ jest zmianą wartości średniej rozkładów składowych rozkładu bimodalnego i Δ < 1/λ2 (obydwie średnie muszą być dodatnie).

Stąd λ’1 = 1/(1/λ1 + Δ) ⇒ λ’2 = 1/(1/λ2 - Δ).

Ten sposób zmiany współczynnika zmienności ogranicza mocno zakres tych zmian do przedziału:   
<1; ~1.75>.

μ

μ1

μ2

μ2’=μ2+Δ

μ1’=μ1-Δ

**1a**. Bardziej ogólnie dla nierównolicznych podzbiorów, tj. %1 ≠ %2 i %1 = 1 - %2, rozsunięcie średnich nie będzie symetryczne, ale jeżeli %1 >> %2 rozsunięcie może być większe:

μ = %1\*μ1 + %2\*μ2 = %1\*μ1’ + %2\*μ2’ = %1 \*(μ1 + Δ1) + %2 \*(μ2 - Δ2)

gdzie: Δ1 = %1/%2 \* Δ2

μ

μ1

μ2

μ2’=μ2+Δ2

μ1’=μ1-Δ1

**Rozwiązanie nr 2** – Przesuwać przez zmianę wartości średniej tylko jeden z rozkładów składowych modyfikując rozmiary podzbiorów: %1’ ≠ %1 i %2’ ≠ %2, tj.:

μ = %1\*μ1 + %2\*μ2 = %1’\***μ1** + %2’\*μ2’ = %1’\***μ1** + %2’\*(μ2 + Δ)

%1’\***μ1** + %2’\*(μ2 + Δ) = %1\*μ1 + %2\*μ2

%1’\***μ1** + (1 - %1’)\*(μ2 + Δ) = %1\*μ1 + %2\*μ2

%1’\*(μ1 - μ2 - Δ) = %1\*μ1 + %2\*μ2 - μ2 - Δ

**%1’ = (%1\*μ1 + (%2 – 1)\* μ2 - Δ)/(μ1 - μ2 - Δ)**

**%2’ = 1 - %1’**

μ

μ2

μ2’=μ2+Δ

μ1’=μ1

**Interesujące wartości parametrów λ - średnia liczba zadań przedkładanych w jednostce czasu i μ - średnia liczba zadań, które może obsłużyć dany system**

Interesującym zakresem powyższych zmiennych jest przypadek, w którym dla fazy krótkich czasów   
λ1 > μ1, a dla długich czasów λ2 < μ2, oczywiście przy zachowaniu warunku, że λśrednie < μśrednie, żeby system pracował stabilnie.