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 zmian w uporządkowaniu czasów między przedkładaniem kolejnych zadań poprzez zmianę liczby dwufazowych cykli**. 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:

$\mu = \%1*\mu_1 + \%2*\mu_2 = \%1'*\mu_1' + \%2'*2\mu_2'$

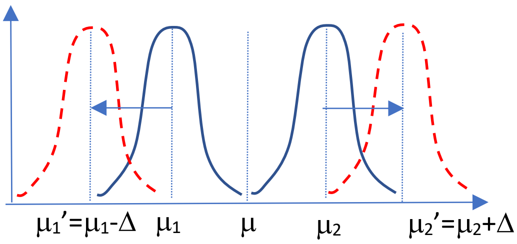
gdzie: wartości %1 i %2 odpowiadają względnej liczebności wartości należących do rozkładów składowych – rozkładu bimodalnego. Dla równolicznych podzbiorów %1 = %2 = 50%:

$\mu = 0{,}5\mu_1 + 0{,}5\mu_2 = 0{,}5\mu_1' + 0{,}5\mu_2' = ((\mu_1 + \Delta) + (\mu_2 - \Delta))/2 = (1/\lambda'_1 + 1/\lambda'_2)/2 = ((1/\lambda_1 + \Delta) + (1/\lambda_2 - \Delta)/2)$

gdzie: Δ jest zmianą wartości średniej rozkładów składowych rozkładu bimodalnego i $\Delta < 1/\lambda_2$ (obydwie średnie muszą być dodatnie).

Stąd $\lambda'_1 = 1/(1/\lambda_1 + \Delta) \Rightarrow \lambda'_2 = 1/(1/\lambda_2 - \Delta)$.

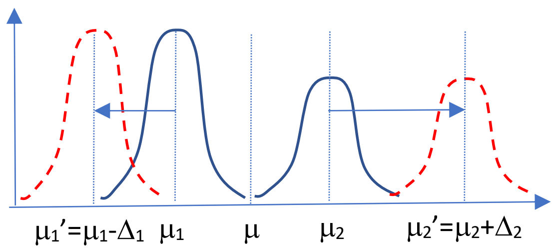
Ten sposób zmiany współczynnika zmienności ogranicza mocno zakres tych zmian do przedziału: <1; ~1.75>.



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

$\mu = \%1*\mu_1 + \%2*\mu_2 = \%1*\mu_1' + \%2*\mu_2' = \%1*(\mu_1 + \Delta_1) + \%2*(\mu_2 - \Delta_2)$

gdzie: $\Delta_1 = \%1/\%2 * \Delta_2$



Rozwiązanie nr 2 – Przesuwać przez zmianę wartości średniej tylko jeden z rozkładów składowych modyfikując rozmiary podzbiorów: %1' $\neq$ %1 i %2' $\neq$ %2, tj.:

$\mu = \%1*\mu_1 + \%2*\mu_2 = \%1'*\boldsymbol{\mu_1} + \%2'*\mu_2' = \%1'*\boldsymbol{\mu_1} + \%2'*(\mu_2 + \Delta)$

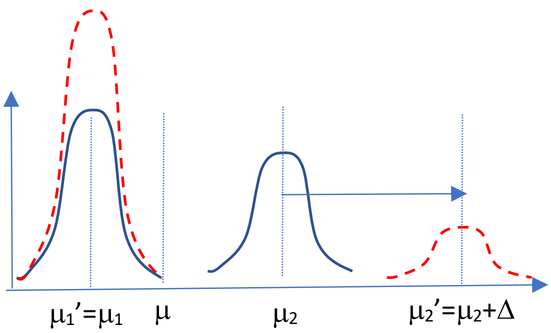
$\%1'*\boldsymbol{\mu_1} + \%2'*(\mu_2 + \Delta) = \%1*\mu_1 + \%2*\mu_2$

$\%1'*\boldsymbol{\mu_1} + (1 - \%1')*(\mu_2 + \Delta) = \%1*\mu_1 + \%2*\mu_2$

$\%1'*(\mu_1 - \mu_2 - \Delta) = \%1*\mu_1 + \%2*\mu_2 - \mu_2 - \Delta$

$\boldsymbol{\%1' = (\%1*\mu_1 + (\%2 - 1)*\mu_2 - \Delta)/(\mu_1 - \mu_2 - \Delta)}$

$\boldsymbol{\%2' = 1 - \%1'}$



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 $\lambda_1 > \mu_1$, a dla długich czasów $\lambda_2 < \mu_2$, oczywiście przy zachowaniu warunku, że $\lambda_{średnie} < \mu_{średnie}$, żeby system pracował stabilnie.