

Raport z testów wydajnościowych systemu ...

ŚRODOWISKO TESTOWE

Testy przeprowadzane były w infrastrukturze klienta na serwerach, które docelowo będą wykorzystywane jako serwery produkcyjne. Jako serwer referencyjny wykorzystano maszynę wirtualną identyczną pod względem parametrów z aktualnym serwerem produkcyjnym.

Konfiguracja serwera referencyjnego

| Warstwa | Web |
|----------------------|-----|
| Liczba potoków vCPU | 12 |
| Pamięć [GB] | 40 |
| Pojemność dysku [GB] | 200 |

Konfiguracja planowanego rozwiązania

| Warstwa | Web | Redis | MariaDB |
|----------------------|-----|-------|---------|
| Liczba potoków vCPU | 12 | 4 | 8 |
| Pamięć [GB] | 8 | 16 | 24 |
| Pojemność dysku [GB] | 200 | 70 | 160 |

W testach rozwiązania referencyjnego obciążenie było generowane z jednego z serwerów wykorzystywanego w nowym rozwiązaniu. W testach nowego rozwiązania obciążenie było generowane z serwera referencyjnego oraz serwera będącego backup'em bazy danych.

ODSTĘPSTWA OD DOCELOWEJ KONFIGURACJI

Infrastruktura testowa nie pozwalała na kierowanie ruchu sieciowego z wykorzystaniem load balancer'a lub innego mechanizmu odpowiedzialnego za dystrybucję zapytań użytkowników do odpowiednich serwerów. Serwery planowanego rozwiązania były obciążane bezpośrednio z maszyn symulujących

zapytania użytkowników. Prezentowane czasy mogą ulec pogorszeniu w przypadku zastosowania, która nie będzie w stanie odpowiednio szybko kierować ruchem sieciowym.

SCENARIUSZ I WYKONANIE TESTÓW

Bazując na danych z Google Analytics dostarczonych przez klienta określono następujące parametry typowej sesji użytkownika:

- Średnia długość sesji: 260 sekund +/- 25%
- Średnia liczba odwiedzanych stron: 5
- Typowa sekwencja akcji użytkownika:
 - 1) Wejście na stronę główną
 - 2) Wejście na stronę losowej kategorii
 - 3) Wyszukiwanie losowego produktu
 - 4) Wejście na stronę losowego produktu
 - 5) Dodanie produktu do koszyka

Test polegał na obciążaniu systemu, które symulowało równoległe odwiedzanie serwisu przez od 200 do 2000 użytkowników. W czasie testu mierzono obciążenie procesora, czas odpowiedzi systemu (czas przesłania wszystkich elementów testowanej strony) oraz liczbę przesłanych stron. Pomiarów czasów odpowiedzi były dokonywane przez generator obciążenia a także przeglądarkę internetową Firefox. Pomiarów dokonane przez generator pozwalają porównać wydajność testowanych rozwiązań i są odporne na zakłócenia spowodowane obciążeniem sieci Internet. Pomiarów takie uzupełniane są wynikami uzyskanymi z rzeczywistej przeglądarki. Na każdym etapie testów dla każdej z testowanych stron dokonywano 5 pomiarów z wykorzystaniem przeglądarki a następnie wyliczano średnią arytmetyczną przedstawioną w raporcie.

W ramach testów porównano wydajność serwera referencyjnego z nowym rozwiązaniem w konfiguracji z:

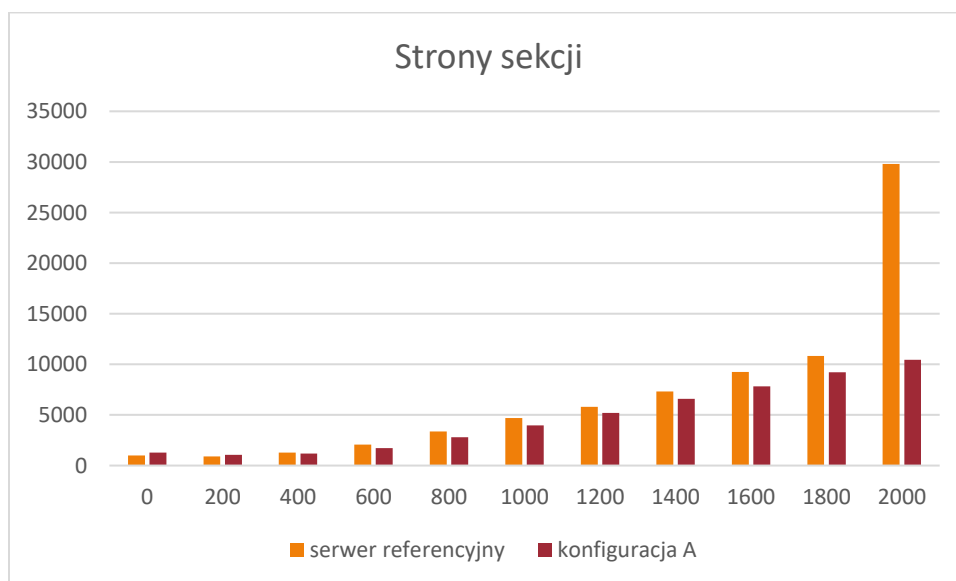
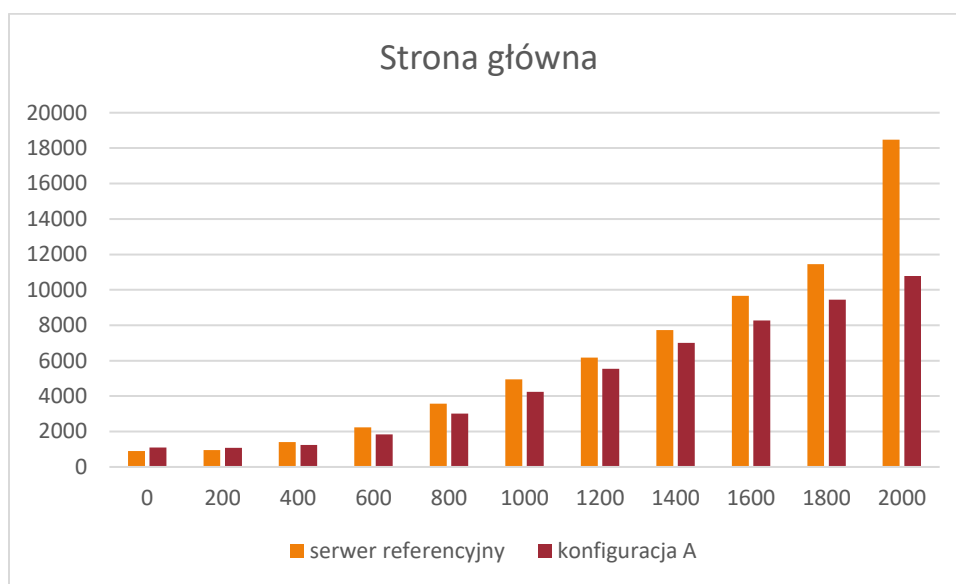
- A. Jednym pracującym serwerem www
- B. Dwoma pracującymi serwerami www

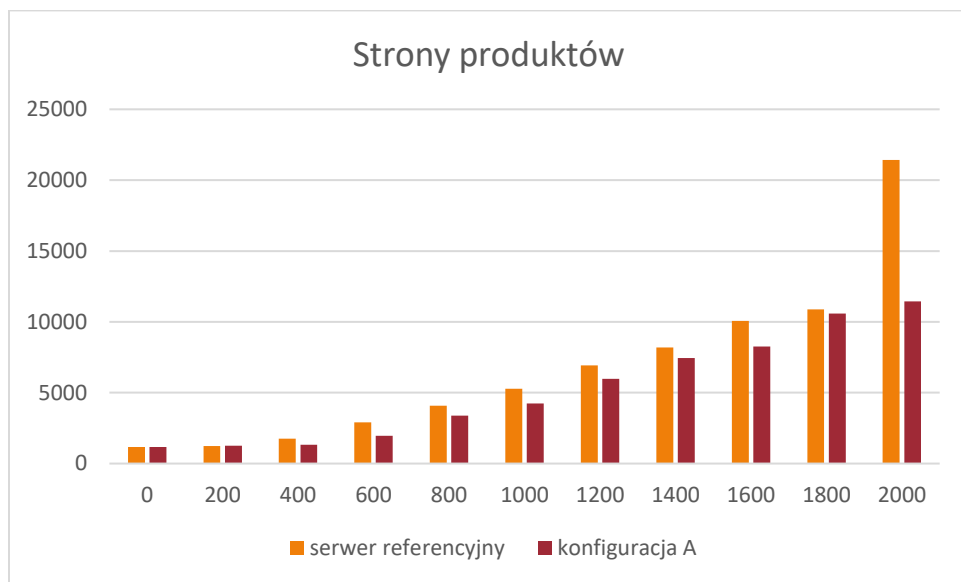
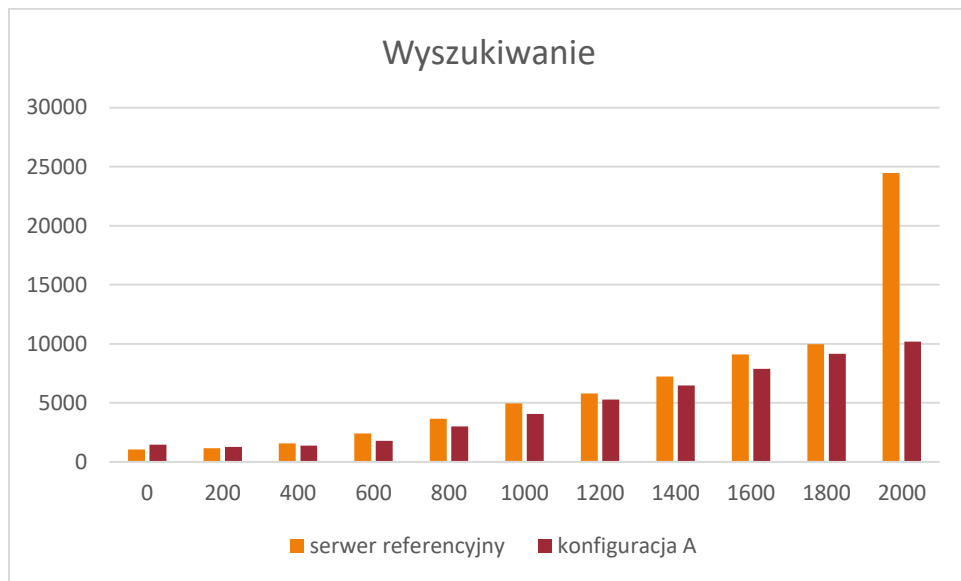
REZULTATY

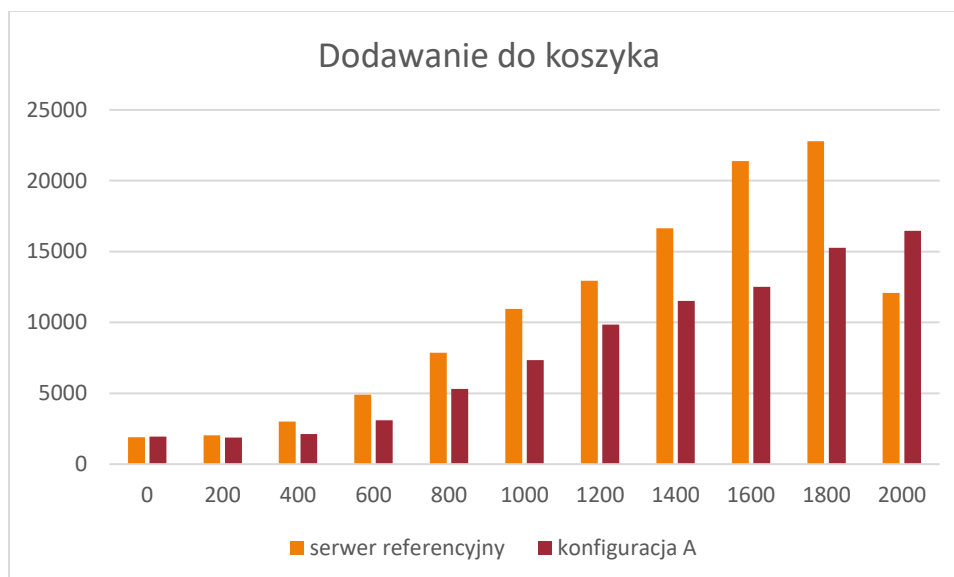
Czasy odpowiedzi podane w raporcie wyrażony jest w milisekundach. Krótszy czas oznacza lepszą wydajność systemu.

Konfiguracja A.

Wyniki uzyskane w czasie testów wskazują, iż referencyjny serwer jest nieznacznie szybszy od nowego rozwiązania, gdy z systemu korzysta mniej niż 400 użytkowników. Wraz ze wzrostem liczby użytkowników korzystających z systemu nowe rozwiązanie generuje i przesyła strony szybciej. Różnica w czasach odpowiedzi zwiększa się wraz ze wzrostem liczby użytkowników. Tendencja ta jest widoczna dla wszystkich testowanych stron.







Krótsze czasy odpowiedzi sprawiły, że nowe rozwiązanie jest w stanie wygenerować ok. 11 stron/sek., podczas gdy rozwiązanie referencyjne tylko ok. 10 stron/sek. Taki wynik pozwala szacować, iż w ciągu godziny serwis w konfiguracji A może wygenerować ok. 40 000.

Serwer referencyjny poza najbardziej wymagającym przypadkiem nie odpowiadał błędnie na zapytania. W przypadku nowego rozwiązania w ramach testów zaobserwowano pojedyncze odpowiedzi z błędami Internal Server Error oraz Service Unavailable.

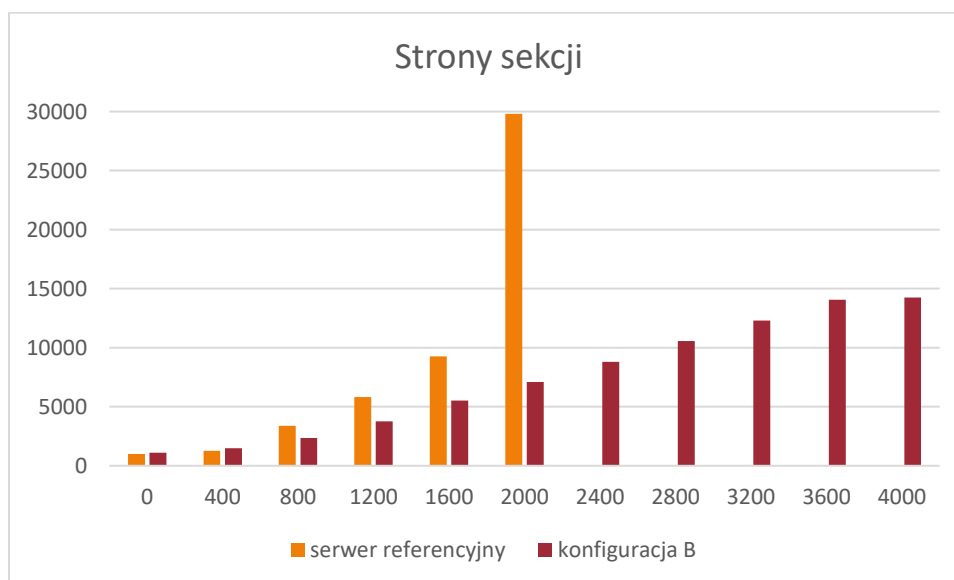
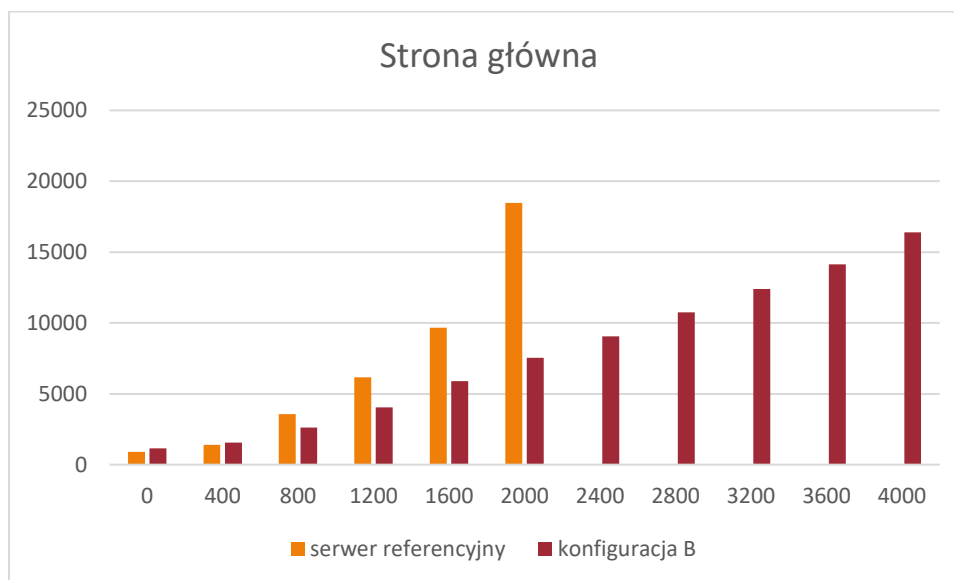
Analizując obciążenie procesora można zauważyć, że obciążenie procesora nie przekracza wartości 70. Oznacza to, iż po przekroczeniu 1000 równoległe korzystających z systemu użytkowników wydajność nowego rozwiązania nie jest ograniczona przez moc procesorów serwera www a inny czynnik.

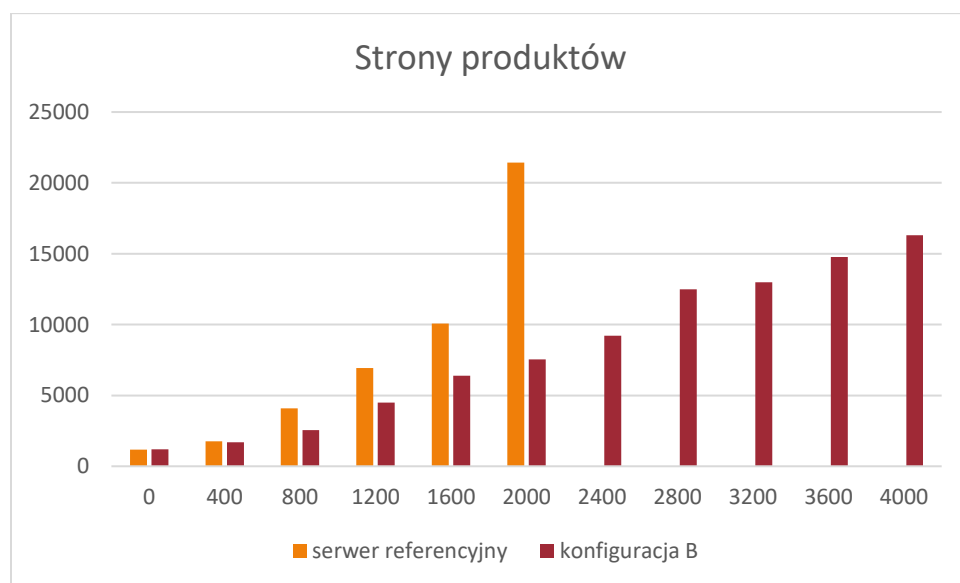
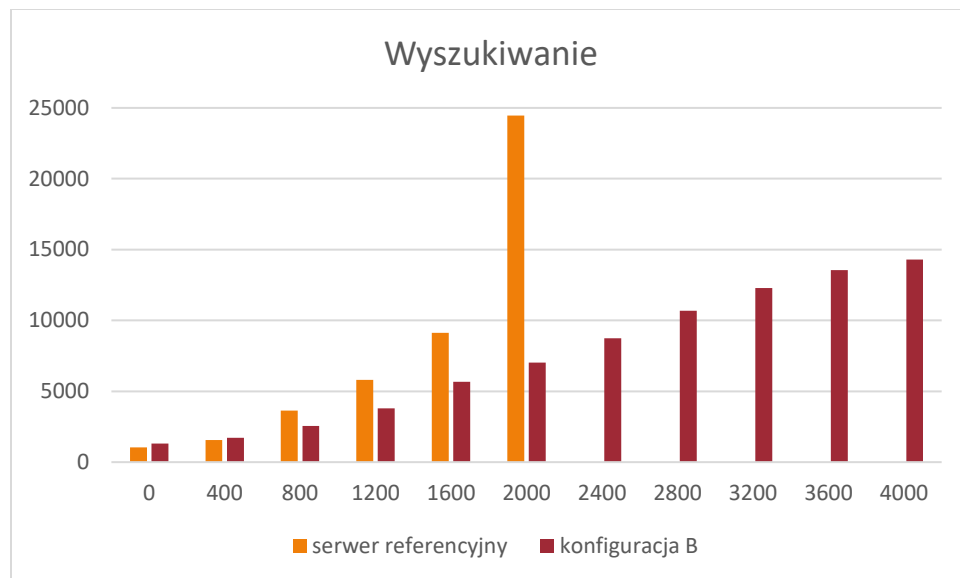
| Liczba użytkowników | Liczba pobranych stron (5 min.) | | Liczba błędów | | load adv. (1min.)* | |
|---------------------|---------------------------------|----------------|---------------------|----------------|---------------------|----------------|
| | serwer referencyjny | konfiguracja A | serwer referencyjny | konfiguracja A | serwer referencyjny | konfiguracja A |
| 200 | 1384 | 1319 | 0 | 1 | 6 | 4 |
| 400 | 2673 | 2561 | 0 | 3 | 12 | 13 |
| 600 | 3167 | 3301 | 0 | 5 | 23 | 25 |
| 800 | 3141 | 3319 | 0 | 4 | 39 | 40 |
| 1000 | 3055 | 3309 | 0 | 4 | 56 | 58 |
| 1200 | 3036 | 3303 | 0 | 7 | 71 | 65 |
| 1400 | 3040 | 3193 | 0 | 4 | 86 | 67 |
| 1600 | 2886 | 3295 | 0 | 5 | 103 | 69 |
| 1800 | 2727 | 3256 | 0 | 13 | 118 | 67 |
| 2000 | 1187 | 3215 | 14 | 6 | b.d. | 69 |

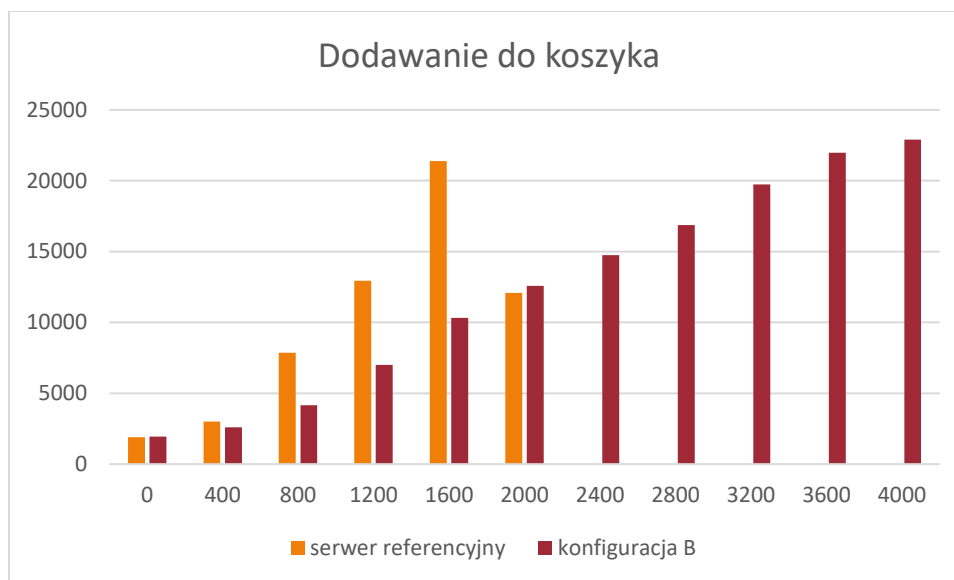
* wartość 1 oznacza całkowite obciążenie jednego z procesorów systemu; wartości większe od liczby procesorów w systemie oznaczają, iż system poddany jest obciążeniu przekraczającemu jego możliwości obliczeniowe

Konfiguracja B.

Zachowanie opisane dla poprzednio omawianej konfiguracji jest jeszcze bardziej widoczne w przypadku konfiguracji z dwoma serwerami www. Większą wydajność nowej architektury serwisu można zaobserwować, gdy liczba użytkowników systemu przekroczy 600 a różnica w czasie odpowiedzi rośnie wraz ze wzrostem liczby użytkowników.





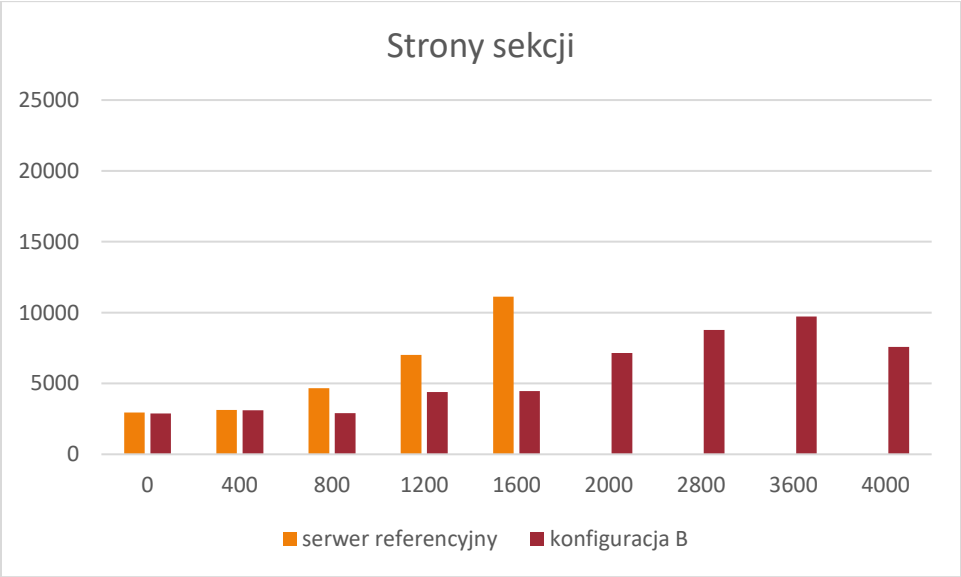
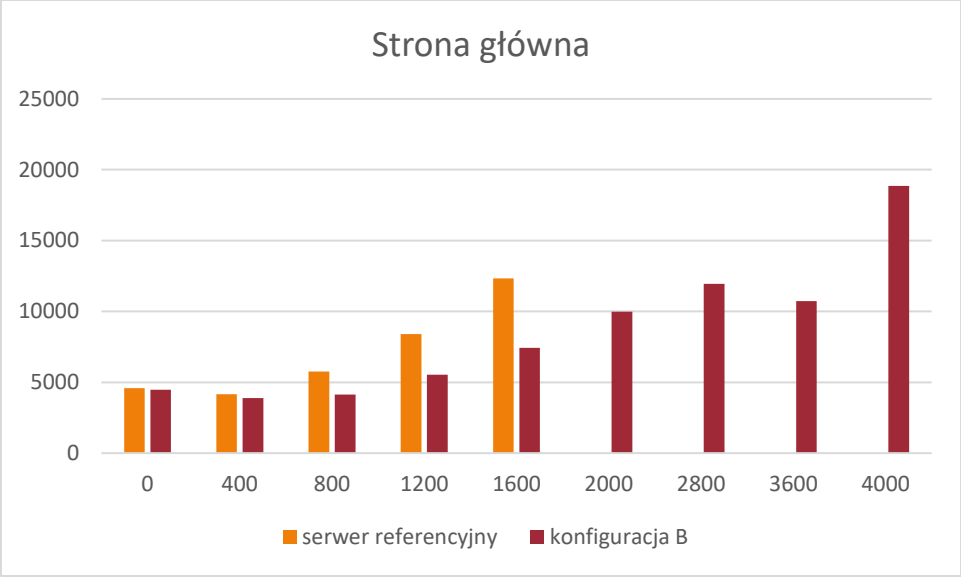


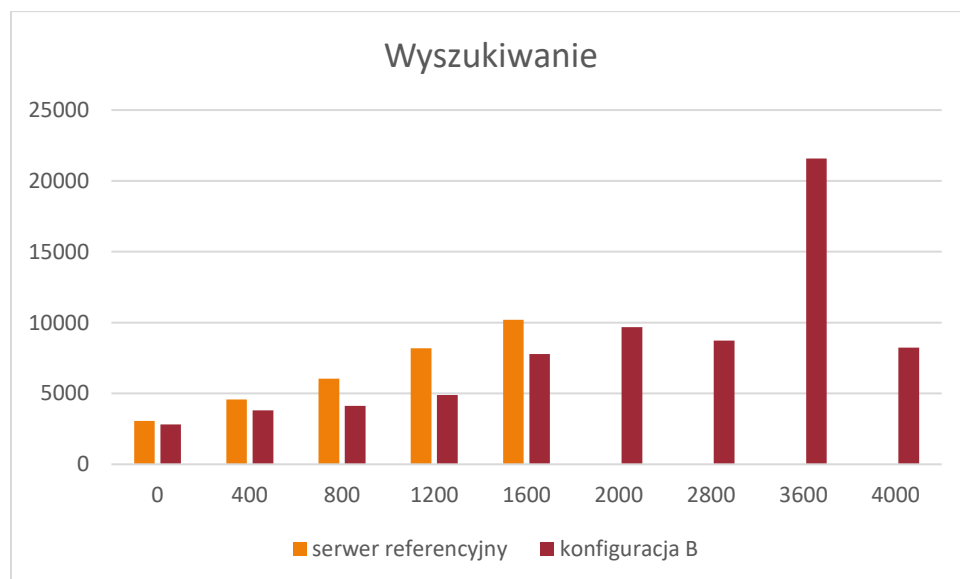
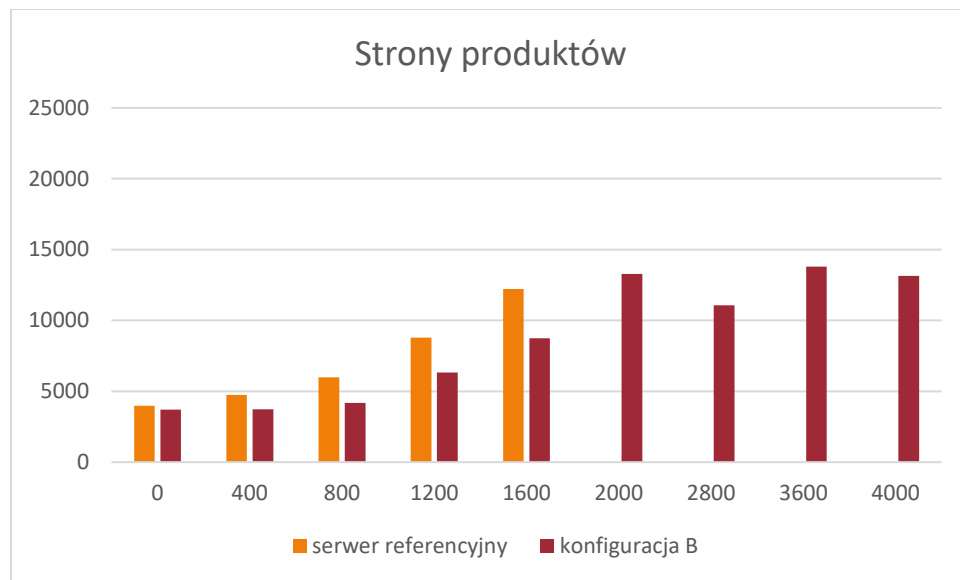
W tej konfiguracji system jest w stanie wygenerować ok. 17 stron/sek. co daje ok 61000 odsłon na godzinę lub ok. 12200 sesji na godzinę.

| Liczba użytkowników | Liczba pobranych stron | | Liczba błędów | | load adv. (1min.)* | |
|---------------------|------------------------|----------------|---------------------|----------------|---------------------|----------------|
| | serwer referencyjny | konfiguracja B | serwer referencyjny | konfiguracja B | serwer referencyjny | konfiguracja B |
| 400 | 2673 | 2522 | 0 | 3 | 12 | 4 |
| 800 | 3141 | 4455 | 0 | 5 | 39 | 13 |
| 1200 | 3036 | 5223 | 0 | 6 | 71 | 25 |
| 1600 | 2886 | 5135 | 0 | 7 | 103 | 40 |
| 2000 | 1187 | 4882 | 14 | 7 | b.d. | 56 |
| 2400 | b.d. | 5058 | b.d. | 7 | b.d. | 65 |
| 2800 | b.d. | 5088 | b.d. | 6 | b.d. | 68 |
| 3200 | b.d. | 4959 | b.d. | 9 | b.d. | 68 |
| 3600 | b.d. | 4887 | b.d. | 23 | b.d. | 68 |
| 4000 | b.d. | 5004 | b.d. | 21 | b.d. | 67 |

* wartość 1 oznacza całkowite obciążenie jednego z procesorów systemu; wartości większe od liczby procesorów w systemie oznaczają, iż system poddany jest obciążeniu przekraczającemu jego możliwości obliczeniowe

Testy pobierania i wyświetlenia strony (dokładnie wygenerowania zdarzenia onLoad) z wykorzystaniem przeglądarki Firefox potwierdziły informacje zebrane w ramach analizy czasów zmierzonych przez generator obciążenia. Przy niedużym obciążeniu systemu rozwiązania podobnie szybko przetwarzają zapytania użytkowników. Różnica w szybkości działania jest widoczna przy większym obciążeniu











PORÓWNANIE CZASÓW WYŚWIETLANIA

Pojedynczy użytkownicy

| 1s | 2sek | 3 sek | 4 sek | 5sek |
|---|---|---|--|---|
|  |  |  |  |  |
|  |  |  |  |  |






400 równolegle korzystających z systemu użytkowników

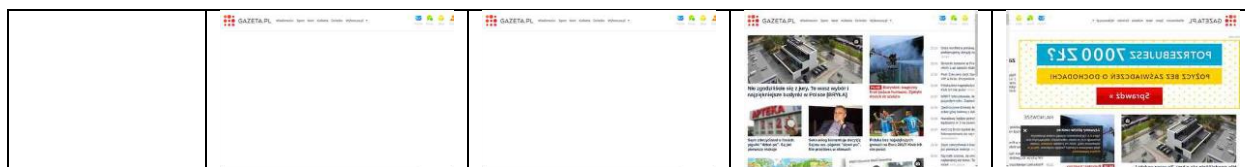
| 1s | 2sek | 3 sek | 4 sek | 5sek |
|--|--|--|---|--|
|  |  |  |  |  |
|  |  |  |  |  |

800 równolegle korzystających z systemu użytkowników

| 1s | 2sek | 3 sek | 4 sek | 5sek |
|---|---|---|--|---|
|  |  |  |  |  |
|  |  |  |  |  |

1200 równolegle korzystających z systemu użytkowników

| 1s | 2sek | 3 sek | 4 sek | 5sek |
|---|---|---|--|---|
|  |  |  |  |  |



PODSUMOWANIE

Proponowane rozwiązanie jest wydajniejsze od rozwiązania referencyjnego a różnica w czasach odpowiedzi systemu rośnie wraz ze wzrostem obciążenia.

W konfiguracji B system pracuje optymalnie (czas wyświetlania stron poniżej 5 sekund a także obciążenie systemu poniżej 24), gdy korzysta z niego równocześnie poniżej 1000 użytkowników. Dla takiej liczby użytkowników system jest w stanie wygenerować w ciągu godziny ponad 57000 stron.

Gdy liczba użytkowników przekroczy granicę 1000 użytkowników system zachowuje się stabilnie jednak czasy oczekiwania na odpowiedź wydłużają i mogą przekraczać nawet 10 sekund.