

VI Seminarium PLOUG
Warszawa
Styczeń 2003

Strojenie systemu Linux pod kątem serwera bazy danych Oracle 9i

Marcin Przepiórowski

1. Wstęp

Strojenie bazy danych Oracle możemy podzielić na kilka etapów. Każdy z nich ma wpływ na ostateczną wydajność naszego środowiska bazodanowego.

Błędy popełnione w poszczególnych etapach strojenia mogą wpływać w niekorzystny sposób na działanie serwera bazy danych. Niektóre z błędów można poprawić stosunkowo niskim nakładem pracy, niektóre z błędów są właściwie nie do naprawienia po wdrożeniu aplikacji.

Błędy które można poprawiać dotyczą tylko i wyłącznie parametrów działania serwera Oracle oraz systemu operacyjnego na którym on działa.

Błędy ciężko naprawialne dotyczą całej reszty, tj:

- analiza potrzeb
- projekt aplikacji
- implementacja aplikacji
- dobór rozwiązań sprzętowo-programowych - zmiana jest możliwa ale często bardzo kosztowna

Jeżeli nasze środowisko zostało dobrze przeanalizowane i zaprojektowane, a faza implementacji aplikacji oraz dobór rozwiązań sprzętowych okazał się dobry, możemy spróbować przyspieszyć działanie naszego serwera poprzez dobór odpowiednich parametrów jego pracy oraz poprzez nastrojenie systemu operacyjnego.

Ten ostatni punkt powyższego wstępu będzie tematem niniejszej pracy. Jak i o ile możemy przyspieszyć pracę serwera Oracle dokonując strojenia parametrów systemu operacyjnego.

Środowisko testowe:

– serwer bazy danych

Linux RedHat 7.3 Pentium III 700 Mhz 256 MB

Dysk SCSI IBM

Oracle 9.2.0.2

- stacja kliencka

Windows 2000 Pentium III 700 Mhz 256 MB

oprogramowanie Quest BenchMark Factor

2. Ustawianie parametrów jądra SO.

W przypadku środowiska opartego o system operacyjny Linux, ważnym zadaniem jest odpowiednie skonfigurowanie a następnie kompilacja jądra SO. W systemach komercyjnych pozostaje nam faza konfiguracji a następnie relinkowanie jądra SO.

Jako punkt wyjścia do porównań przyjąłem standardową instalację Linux-a RedHat 7.3 wraz z kerneliem 2.4.18-3. Jest to jądro patchowane przez firmę RedHat pod kątem ich dystrybucji.

TPC-C	13.039
Avg Time	0.211 s
Ładowanie danych SQLLDR (90000 rekordów)	11.281 s
Ładowanie tabeli do testów TPC-C (100000)	5 m 21 s

Jądro dostarczane standardowo przez firmę RedHat skompilowane jest z wieloma modułami dołączanymi dynamicznie do jądra w miarę potrzeb.

Ta modularność wpływa niekorzystnie na wydajność działania systemu operacyjnego.

Kolejny test polegał na skompilowaniu jądra bez modułów, przy okazji zmieniając wersje z 2.4.18-3 (RedHat) na wersję 2.4.19 (www.kernel.org). Zmiany tej dokonano, aby uniezależnić testy od danej dystrybucji i zmian jakie dokonywane są przez jej producentów w jądrze SO.

TPC-C	13.442
Avg Time	0.19 s
Ładowanie danych SQLLDR (90000 rekordów)	10.608
Ładowanie tabeli do testów TPC-C (100000)	5 m 16 s

Kolejny test był wykonany na najnowszej dostępnej w chwili wykonywania testów wersji jądra Linux-a czyli na wersji 2.4.20

TPC-C	13.973
Avg Time	0.183 s
Ładowanie danych SQLLDR (90000 rekordów)	9.747
Ładowanie tabeli do testów TPC-C (100000)	5 m 17 s

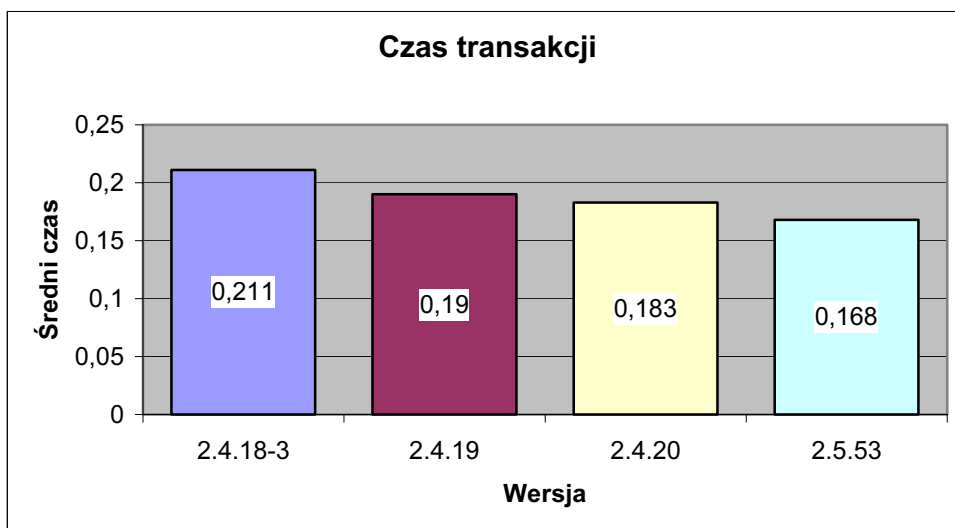
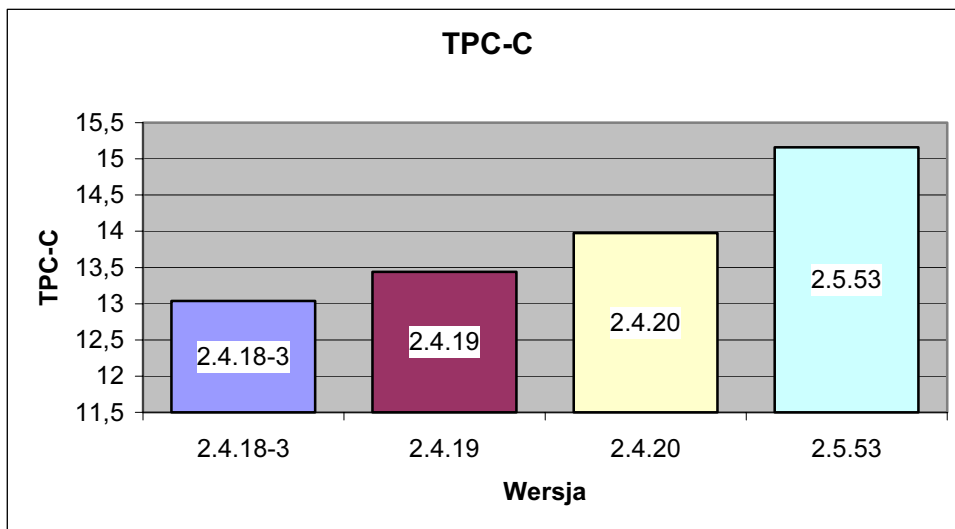
Nowszy kernel zwiększa wydajność a również niezawodność działania, poprzez poprawienie błędów znalezionych w wersjach poprzednich. Idąc za ciosem, do następnej próby został wykorzystany najnowsza niestabilna wersja jądra 2.5.53.

TPC-C	15.159
Avg Time	0.168
Ładowanie danych SQLLDR (90000 rekordów)	10.321
Ładowanie tabeli do testów TPC-C (100000)	5 m 19 s

Przyrost jest dość znaczny, ale w serii 2.5.x dokonano dużo zmian i ulepszeń w stosunku do wersji 2.4.x. Pojawienie się serii 2.6.x będącej stabilną (produkcyjną) wersją linii 2.5.x będzie dużym krokiem naprzód w budowaniu wydajnych i tanich rozwiązań bazodanowych.

Podsumowanie:

	2.4.18-3	2.4.19	2.4.20	2.5.53
TPC-C	13.039	13.442 + 3,09 %	13.973 + 7,16 %	15.159 + 16,25 %
Avg Time	0.211 s	0.19 s - 9,95 %	0.183 s - 13,27 %	0.168 - 20,37 %
Ładowanie SQLDR	11.281 s	10.608 - 5,96 %	9.747 - 13,59 %	10.321 - 8,50 %
Ładowanie	5 m 21 s	5 m 17 s	5 m 16 s	5 m 19 s



3. Strojenie parametrów jądra

Strojenie parametrów jądra możemy podzielić na dwa etapy:

1. strojenie parametrów związanych z pamięcią dzieloną oraz semaforami
2. strojenie działania mechanizmów buforowania dostępu do plików

Pierwszy etap musi być wykonany w większości systemów operacyjnych aby dostosować system do wymagań serwera Oracle pod kątem pamięci współdzielonej i liczby semaforów dostępnych w systemie. Niedostosowanie tych parametrów może wpływać na możliwości konfigurowania instancji Oracle, przede wszystkim na liczbę procesów (parametr PROCESS) oraz na maksymalny rozmiar pamięci SGA.

Etap drugi pozwala na dostosowanie (polepszenie) działania systemu buforowania plików do charakteru pracy bazy danych Oracle.

TPC-C	14.903
Avg Time	0.168
Ładowanie danych SQLLDR (90000 rekordów)	12.302
Ładowanie tabeli do testów TPC-C (100000)	5 m 10 s

4. Ustawianie parametrów filesystemu

Wszystkie dotychczasowe testy były prowadzone w oparciu o system plików „ext2” – standardowy system plików Linux-a, bez opcji journalingu.

W większości systemów operacyjnych, każda modyfikacji pliku, powoduje zmianę parametru „access time” czyli czas ostatniego dostępu do pliku.

Modyfikacja tego parametru, wymaga czasu i zasobów procesora oraz urządzenia dyskowego. W większości systemów można wyłączyć tę właściwość systemu plików uzyskując przy tym polepszenie wydajności I/O, a co za tym idzie polepszenie działania serwera Oracle.

TPC-C	15.082
Avg Time	0.165
Ładowanie danych SQLLDR (90000 rekordów)	12.766
Ładowanie tabeli do testów TPC-C (100000)	5 m 51 s

Kolejna rzecz do właściwy dobór rozmiaru bloku w systemie plików. Blok ten powinien być mniejszy lub równy blokowi bazy danych Oracle. Dzięki temu nie na nadmiarowych odczytywanych bajtów przy dostępie do konkretnego bloku bazodanowego. Wszystkie testy były przeprowadzane z blokiem o rozmiarze 4 kB. Dla porównania został wykonany test na bloku o rozmiarze 2 kB.

TPC-C	14.908
Avg Time	0.167
Ładowanie danych SQLLDR (90000 rekordów)	9.89 s
Ładowanie tabeli do testów TPC-C (100000)	5 m 01 s

Poniższe zestawienie obrazuje nam przyrost wydajności serwera bazy danych w zależności od modyfikacji parametrów jądra i parametrów systemu plików.

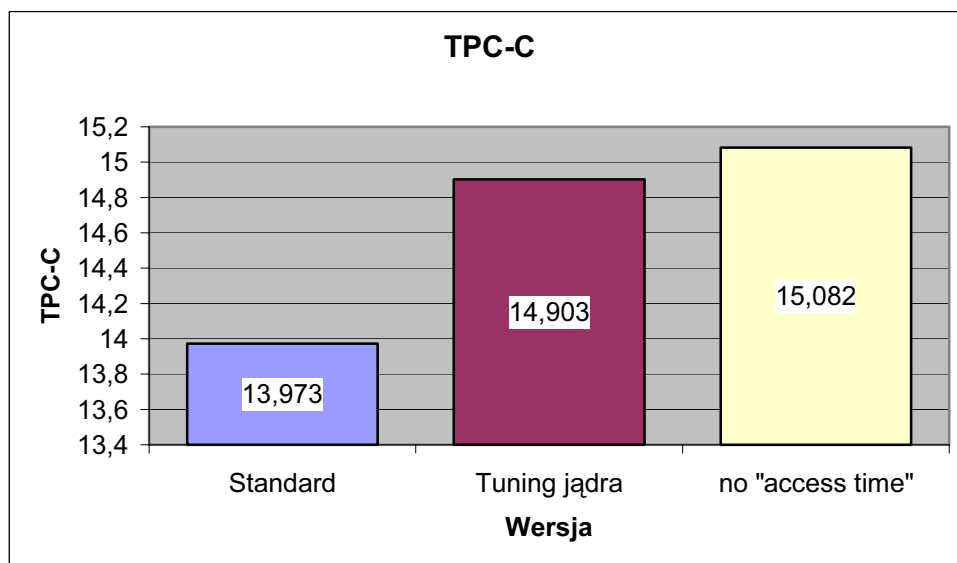
	Standard	Zmiana buforowanie i pamięci współdzielonej	Brak modyfikacji „access time”
TPC-C	13.973	14.903	15.082
Avg Time	0.183 s	0.168	0.165
Ładowanie SQLLDR	9.747	12.302	12.766
Ładowanie	5 m 16 s	5 m 10 s	5 m 51 s

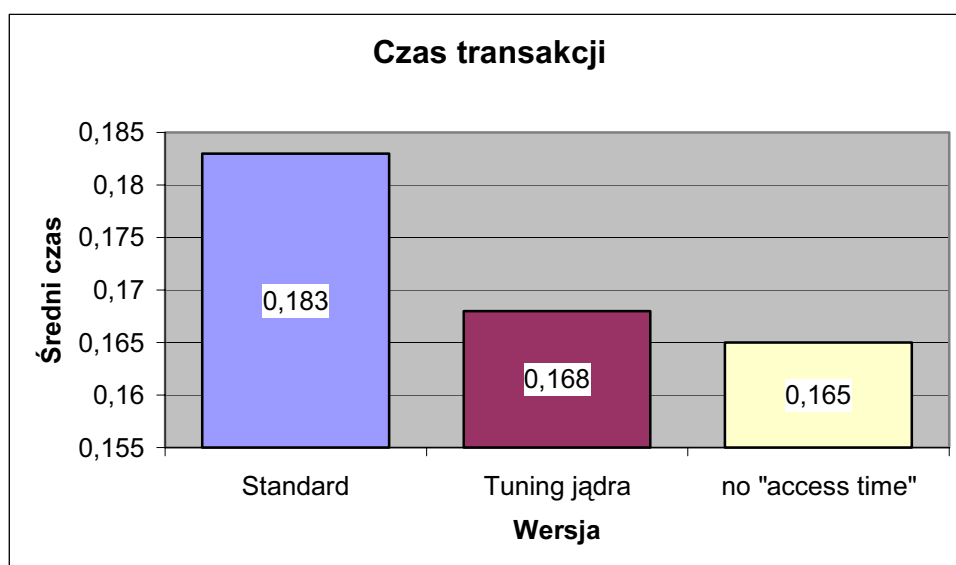
Porównanie wydajności przy różnych rozmiarach bloku systemu operacyjnego.

Większy blok – lepsza wydajność TPC-C

Mniejszy blok – szybsze ładowanie – głównie z uwagi na częste zapisy bufora redo na dysk – mniejszy blok SO szybsze zapisy.

	Blok 4 kB	Blok 2 kB
TPC-C	15.082	14.908
Avg Time	0.165	0.167
Ładowanie SQLLDR	12.766	9.89 s
Ładowanie	5 m 51 s	5 m 01 s





5. Wybór optymalnego filesystemu

Ponieważ w systemie Linux dostępne są różne systemy plików, zostały one porównane pod kątem ich wydajności przy używaniu bazy danych Oracle. Punktem wyjściowym był standardowy system plików Linux-a czyli „ext2”.

TPC-C	15.082
Avg Time	0.165
Ładowanie danych SQLLDR (90000 rekordów)	12.766
Ładowanie tabeli do testów TPC-C (100000)	5 m 51 s

Kolejny system plików to rozwinięcie „ext2” o możliwości journalingu, czyli system „ext3”.

TPC-C	14.593
Avg Time	0.173
Ładowanie danych SQLLDR (90000 rekordów)	13.370
Ładowanie tabeli do testów TPC-C (100000)	5 m 20 s

Następnie przetestowany został system „resier fs”

TPC-C	14.961
Avg Time	0.166
Ładowanie danych SQLLDR (90000 rekordów)	14.804
Ładowanie tabeli do testów TPC-C (100000)	6 m 49 s

7Następne dwa systemy plików zostały zaimplementowane na innych systemach operacyjnych i przeniesione na Linux-a. Pierwszy z nich jest produktem firmy IBM i nazywa się „JFS”.

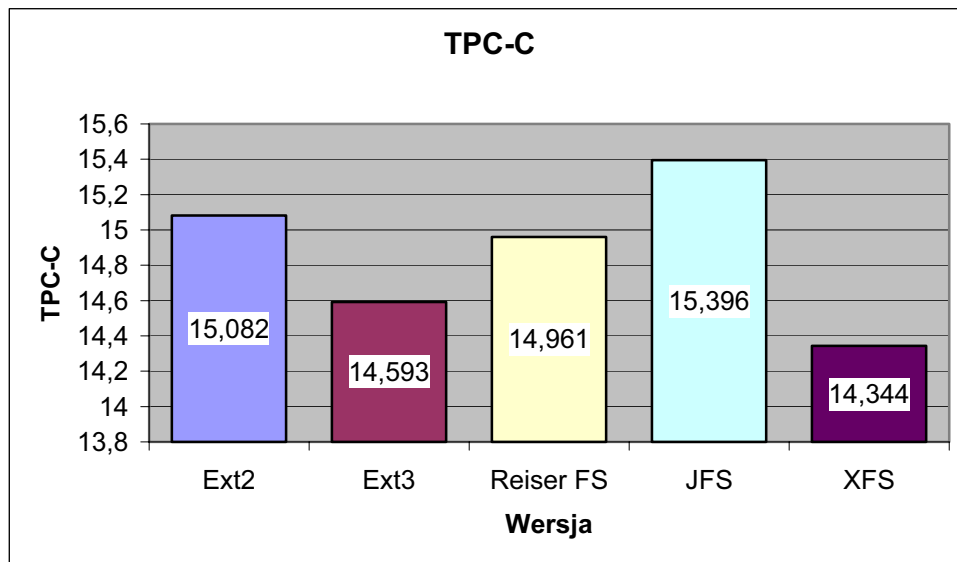
TPC-C	15.396
Avg Time	0.159
Ładowanie danych SQLLDR (90000 rekordów)	13.046
Ładowanie tabeli do testów TPC-C (100000)	5 m 01 s

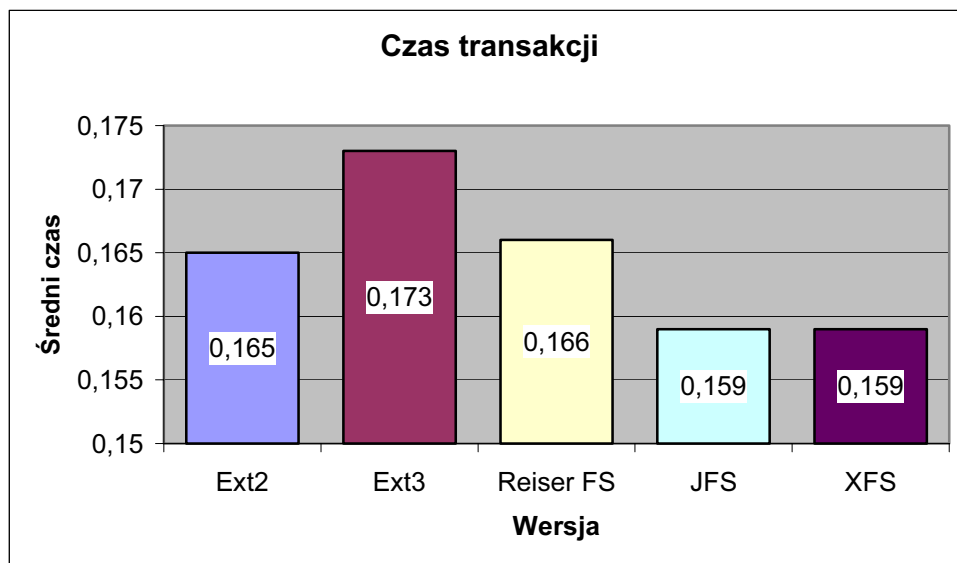
Drugi z nich to „XFS” firmy SGI, używany w systemie IRIX.

TPC-C	14.344
Avg Time	0.159
Ładowanie danych SQLLDR (90000 rekordów)	11.292
Ładowanie tabeli do testów TPC-C (100000)	5 m 24 s

Zestawienie

	Ext2	Ext3	Reiser FS	JFS	XFS
TPC-C	15.082	14.593 - 3,24 %	14.961 - 0.8 %	15.396 + 2,08 %	14.344 - 4,89 %
Avg Time	0.165	0.173 + 4,84 %	0.166 + 0,6 %	0.159 - 3,63 %	0.159 - 3,63 %
Ładowanie SQLLDR	12.766	13.370	14.804	13.046	11.292
Ładowanie tabeli TPC-C	5 m 51 s	5 m 20 s	6 m 49 s	5 m 01 s	5 m 24 s





6. Pamięć SGA i „swap”.

W przypadku, gdy kończy nam się pamięć w systemie operacyjnym i zaczyna się proces „swapowania”, wydajność bazy danych Oracle spada dość szybko. Jeżeli nie możemy zwiększyć zasobów pamięciowych serwera można spróbować wymusić zablokowanie całej pamięci SGA w pamięci RAM, za pomocą parametru „lock_sga=true”.

Bez swapowania

TPC-C	15.396
Avg Time	0.159

Ze swapowaniem

	bez parametru	z parametrem
TPC-C	14.320	14.981
Avg Time	0.180	0.165

Oczywiście wydajność spadała w porównaniu do ostatniego wyniku, nawet przy zabokowaniu pamięci SGA, ponieważ procesy użytkowników już korzystają z która została umieszczona w obszarze swap.

7. Podsumowanie

Strojenie systemu operacyjnego powinno być ostatnim etapem strojenia środowiska bazodanowego opartego o bazę Oracle i nie może zastępować poprzednich etapów.

Przyspieszenie, jakie możemy uzyskać jest stosunkowo niewielkie, w porównaniu do zysku z innych etapów, ale etap ten nie powinien być pomijany, ponieważ pozwala on zgrać wszystkie elementy środowiska w jedną całość.