



Co oferuje SAS Scalable Performance Data Server?

SAS Scalable Performance Data Server jest zoptymalizowany na dostarczenie podzbiorów informacji, które trzeba szybko wydobyć z wielkich wolumenów danych przedsiębiorstwa realizując dowolne żądanie.

Dlaczego SAS Scalable Performance Data Server jest tak istotny?

Możliwości SAS Scalable Performance Data Servera zapewniają aplikacjom typu business intelligence i aplikacjom analitycznym stabilną i wysoką wydajność oraz to, że poszczególne etapy procesu ETL zmieszczą się w przeznaczonych im okienkach czasowych pomimo stałego wzrostu wolumenu danych w przedsiębiorstwie.

Dla kogo przeznaczony jest SAS Scalable Performance Data Server?

SAS Scalable Performance Data Server został zaprojektowany z myślą o menadżerach IT zarządzających wielkimi wolumenami danych, którzy są zainteresowani przyspieszeniem działania aplikacji typu business intelligence i aplikacji analitycznych lub zmniejszeniem czasów: pozyskania, transformacji i ładowania danych do hurtowni niezależnie od rozmiaru danych przechowywanych w przedsiębiorstwie.

SAS® Scalable Performance Data Server®

Repozytorium zoptymalizowane do szybkiego i bezzwłocznego dostarczania podzbiorów danych.

Ci, którzy podczas podejmowania decyzji, polegają na rozwiązaniach business intelligence i aplikacjach analitycznych, poszukują jednego - precyzyjnych odpowiedzi wtedy, gdy są potrzebne. Niestety, dane potrzebne do tych odpowiedzi są często ukryte w wielkich wolumenach danych.

Zorientowane na przetwarzanie transakcji relacyjne bazy danych od dawna skupiają się bardziej na gromadzeniu danych niż na ich dostarczeniu dla raportów i analiz. Nie były też one projektowane pod kątem wyszukiwania informacji, tworzenia raportów i analiz na wielką skalę. Z tego powodu, przetwarzania tego typu wykonywane na relacyjnych bazach danych mogą prowadzić do zmniejszenia wydajności systemu, zużycia dużej ilości jego zasobów i do kolizji z działalnością operacyjną. Aby dostosować się do oczekiwań użytkowników podzbiory danych często muszą być przenoszone na inne maszyny, co oprócz opóźnień w tworzeniu raportów powoduje trudności w utrzymaniu spójnego obrazu przedsiębiorstwa.

Wraz z oprogramowaniem SAS Scalable Performance Data Server użytkownik otrzymuje repozytorium (storage system), które jest zoptymalizowane zarówno na szybkie ładowanie danych jak i szybki odczyt podzbiorów z wielkich wolumenów danych. Raportowanie dla biznesu, a również wielkiej skali aplikacje analityczne skorzystają z szybkiego dostępu do danych szczegółowych przechowywanych przez ten serwer. Zapewnia on aplikacjom stabilną wydajność oraz to, że procesy ETL-owe nie przekroczą okienek czasowych nawet przy stale wzrastającej ilości danych.

SAS Scalable Performance Data Server osiąga swoją skalowalność poprzez przetwarzanie równoległe, partycjonowanie tabel, bitmapowy indeks hybrydowy z B-Tree, optymalizowanie zapytań oraz eliminację narzutów związanych z typowym przetwarzaniem transakcji. Gdy wydobywanie danych, transformacja i załadowanie zajmują mniej czasu, wtedy raporty są generowane szybciej a podejmujący decyzje mogą wcześniej otrzymać odpowiedzi, których potrzebują.

Kluczowe korzyści

- ❖ **Znaczące przyspieszenie wydobywania podzbiorów informacji.** Większość aplikacji business intelli-

gence i aplikacji analitycznych wymaga jedynie podzbiorów informacji znajdujących się w wielkich repozytoriach danych przedsiębiorstwa. SAS Scalable Performance Data Server używa technologii przetwarzania równoległego i wykorzystuje możliwości współczesnego sprzętu komputerowego, oraz redukuje narzuty takie jak dwu-fazowy commit, który jest związany z systemami transakcyjnymi. To znacząco zmniejsza czas odszukania i dostarczenia danych.

- ❖ **Zoptymalizowana wydajność narzędzi business intelligence i aplikacji analitycznych.** Dzięki odseparowaniu środowiska transakcyjnego od środowiska aplikacji analitycznych i business intelligence, każde z nich można zoptymalizować tak, aby najlepiej spełniało wymagania poszczególnych zadań. Otrzymujemy lepszą wydajność wykonywania raportów i analiz jednocześnie zmniejszając obciążenia systemów wspomagających działalność operacyjną
- ❖ **Zmniejszenie czasów pozyskania, transformacji i ładowania danych.** Zintegrowany proces ETL przenosi dane z systemów transakcyjnych do SAS Scalable Performance Data Servera wykorzystując równoległe działające repozytorium i wydajne mechanizmy indeksowania. To ogranicza czas pozyskania, transformacji i ładowania danych pozwalając, aby dane i indeksy były szybko odświeżane niezależnie od ich rozmiarów. Wykorzystanie zintegrowanych metadanych zapewnia spójność danych i zmniejsza czas potrzebny na utrzymywanie wielkich baz danych.

Opis produktu

Posiadając SAS Scalable Performance Data Server departamenty IT mogą zapewnić wielu równocześnie pracującym użytkownikom szybki i bezpieczny dostęp do podzbiorów informacji wybieranych z potężnych pamięci masowych przedsiębiorstwa. Możliwości serwera pozwalające na wykorzystanie skalowalności istniejącego sprzętu (w zależności od rozmiaru problemu) do odświeżania wielkich zbiorów danych w określonym oknie czasowym dają maksymalny zwrot z inwestycji. SAS Scalable Performance Data Server sprawdza się najlepiej tam, gdzie mamy do czynienia z wielkimi zbiorami danych.

Maksymalny rozmiar danych w hurtowni i czas spędzony na ich przetwarzanie są limitowane przede wszystkim przez dostępny sprzęt. Serwer został zaprojektowany tak, aby skalować się ze sprzętem (scale up) wraz z dodawaniem do niego dodatkowych komponentów takich jak: procesory, pamięć operacyjna, podsystemy I/O.

Aby zaoferować najszybszy dostęp do wielkich zbiorów danych SAS Scalable Performance Data Server oparty jest o wielowątkowy motor I/O. Motor ten jednocześnie uruchamia wiele niezależnych wątków (tzw. procesów lekkich lub LWP's), przeszukujących równoległe dużą tabelę celem odnalezienia wierszy zawierających oczekiwaną informację. To skaluje (skraca) czas potrzebny na znalezienie informacji do ułamka czasu, który potrzebowałibyśmy w pojedynczym jedno-wątkowym przeszukiwaniu sekwencyjnym.

Partycjonowanie i przetwarzanie równoległe zapewnia skalowalność i wydajność Skalowalne I/O

SAS Scalable Performance Data Server przyspiesza przetwarzania dużych zbiorów danych poprzez ich partycjonowanie (rozłożenie) po wielu dyskach i poprzez jednoczesne wykorzystanie wielu kanałów I/O. Dzięki temu SPD Serwer może wykonywać równoległe wiele funkcji I/O pracując jednocześnie z wieloma partycjami danych. SPD Serwer został zaprojektowany tak, aby wykorzystywać wszystkie zasoby dostępne na komputerze w celu uzyskania maksymalnej skalowalności. Dlatego największe korzyści są osiągnięte wtedy, gdy sprzęt wyposażony jest w wiele procesorów, odpowiednią liczbę kanałów I/O i urządzeń dyskowych oraz tam gdzie przetwarzane wolumeny danych są naprawdę wielkie.

Niektóre istniejące w praktyce implementacje SAS Scalable Performance Data Server przetwarzają dane wielkości wielu terabajtów przy pojedynczych tablicach przekraczających 500 GB. SAS Scalable Performance Data Server zademonstrował swoją skalowalność obsługując tablice z miliardami obserwacji a został zaprojek-

nowany dla przetwarzania danych przy użyciu przestrzeni adresowej rzędu petabajtów, po to, aby dobrze wspierać wielkie hurtownie danych.

Możliwości równoległego Group By

Możliwości Parallel Group By (PGB) to równoległe przetwarzanie wykorzystujące procesory i równoległe I/O dla wybranych rodzajów agregacji. Wspierane są takie funkcje jak: COUNT, FREQ, N, AVG, MEAN, MAX, MIN, NMISS, STD, SUM, VAR i DISTINCT. W nowej wersji wspierane są także tzw. „aliasy”.

Możliwości nowego PGB są bezpośrednio zintegrowane z motorem SAS SQL, przez co zwiększa się liczba sytuacji, w których PGB może być zastosowany.

Wielowątkowe przetwarzanie warunków WHERE

Inteligentny podsystem planowania przetwarzania warunków WHERE SPD Serwera wybierze najszybszą strategię odnalezienia dowolnego podzbioru danych, niezależnie od liczby indeksów nałożonych na zmienne w warunku. Największe korzyści ze skalowalności przynosi ocena skomplikowanych klauzul zawierających złożenie wielu warunków, a także zastosowanie strategii full table scan, np. gdy tabela nie ma zdefiniowanych indeksów. W takich sytuacjach SPD Serwer powołuje wątki do przetwarzania różnych fragmentów predykatu lub całej tablicy (przy full table scan) równocześnie zapewniając dużą przepustowość I/O i skalowalność.

Mechanizmy szacowania kosztów (WHERE planner costing) oceniają efektywność różnych strategii przetwarzania zapytania w celu wybrania tych najmniej obciążających zasoby. Wybór obejmuje m.in. decyzje czy i których indeksów używać a bazuje na analizie zagęszczenia i rozkładzie wartości w indeksie.

Unikalna technologia indeksu ogranicza zajętość dysku i przyspiesza zapytania

SAS Scalable Performance Data Server wykorzystuje schemat indeksu, który

łączy najlepsze cechy indeksów typu B-tree z indeksami bitmapowymi. SAS Scalable Performance Data Server analizuje każdą tabelę, automatycznie określając najlepszy sposób indeksowania dla każdego segmentu danych. Wynikiem jest większa wydajność, szybsze łączenie tablic i zmniejszone wymagania na pamięć masową.

Równoległe tworzenie i modyfikacje indeksów

SAS Scalable Performance Data Server może tworzyć i odświeżać wiele indeksów jednej tabeli równoległe również wtedy, gdy tabela jest ładowana przy użyciu procedury APPEND. SPD Serwer buduje wiele indeksów podczas pojedynczego odczytu tabeli, co bardzo przyspiesza czas ich tworzenia.

Index statistics table

SAS Scalable Performance Data Server utrzymuje tablicę ze statystykami opisującymi indeksy, aby na podstawie kosztu ich użycia, lepiej ocenić, kiedy je wykorzystywać. Obejmuje to również sytuacje związane z wyborem strategii łączenia tabel (index join, merge join, hash join) w zapytaniach SQL.

Zoptymalizowana wydajność zapytań Optymalizacja SQL Plannera

Nowe mechanizmy optymalizacji podsystemu SQL Planner SPD Serwera zwiększają wydajność najbardziej powszechnie spotykanych zapytań zadawanych z aplikacji typu business intelligence i aplikacji analitycznych.

Zoptymalizowana obsługa tzw. „correlated queries” – w przezroczyście sposób rekonstruuje i przekształca zagnieżdżony kod SQL, tworząc „w locie” zbiory tymczasowe, co przyspiesza złożone przetwarzania a jednocześnie zachowuje ich logikę biznesową.

Wydajne sortowanie przy instrukcjach BY

Sortowanie tablic jest operacją stosunkowo często spotykaną. SAS Scalable Performance Data Server zwiększa wydajność sortowania na dwa sposoby.

Po pierwsze w przypadkach, gdy występuje warunek WHERE lub instrukcja KEEP, które ograniczają zbiór wejściowy, sortowany będzie tylko wybrany podzbiór powodując, że czas operacji będzie zależny od wielkości podzbioru a nie wielkości całego zbioru. W takim przypadku zużycie zasobów systemowych może być bardzo ograniczone. Dodatkowo instrukcje BY, bez względu na to gdzie występują, nie wymagają wcześniejszego fizycznego sortowania zbioru, gdyż odbędzie się ono niejako „w locie”.

Po drugie 64-bitowa (wide-mode) wersja SAS Scalable Performance Data Server



może stworzyć w pamięci operacyjnej pojedynczy obszar danych sortowania (sort bin) większy niż 2GB o ile dysponujemy wystarczającą jej ilością. Takie rozwiązanie ogranicza wykorzystanie pamięci masowej i prowadzi do znaczącego skrócenia czasu sortowania.

Wsparcie dla SQL passthrough

Procedura SQL, dzięki funkcjonalności pass-through, daje możliwość przestania specyficznego dla SPD Serwera kodu SQL bezpośredniego do wykonania przez serwer. Wprawdzie największe korzyści uzyskiwane są przy kodzie specyficznym dla SAS Scalable Performance Data Server, ale można też wykorzystać „pass-through” do uruchamiania wielu niezależnych zapytań SQL-owych równoległe.

Automatic aging rollout

Ta funkcjonalność to m.in. ułatwienie ładowania i usuwania danych z wielkich tabel przez ich zorganizowanie w fizycznie odrębne fragmenty zdefiniowane przez stempel czasowy. Scalable Performance Data Server używa wirtualnej tablicy nazywanej klastrem zawierającej określoną ilość tzw. slotów, z których każdy przechowuje dane dotyczące ustalonego przedziału czasu. Kiedy wszystkie sloty są wypełnione i dodajemy nowe dane, najstarszy slot jest automatycznie kasowany, aby zwolnić miejsce nowym danym. W ten sposób tabela może przechowywać automatycznie jedynie najnowsze 12 miesięcy danych lub 4 lata historii, itp.

Wsparcie dla 64-bitów (Wide-Mode)

Wsparcie dla systemów 64-bitowych umożliwi serwerowi alokowanie i adresowanie ponad 2 GB pamięci operacyjnej na raz, co przyspiesza przetwarzania dużych zbiorów danych przy operacjach takich jak np. sortowanie.

Konsolidacja SPDSSNET-a

W poprzednich wersjach produktu, SPDSSNET był mechanizmem zapewniającym wydajny dostęp do SAS Scalable Performance Data Serwera aplikacjom typu desktop działającym pod kontrolą Windows. Wraz z SAS SPD Serwer 9.1 połączenie ODBC jest realizowane poprzez standardowy port zdefiniowany dla SPDS-a (tj. przez Name Server).

Zaawansowana implementacja bezpieczeństwa

Funkcje bezpieczeństwa SAS Scalable Performance Data Server pozwalają administratorom kontrolować uprawnienia do odczytu, zapisu i modyfikacji tablic na poziomie pojedynczego użytkownika lub grupy.

Główne funkcje

Partycjonowanie danych i przetwarzanie równoległe

- ▶ Partycjonowanie danych na wiele procesorów, kanałów I/O i dysków
- ▶ Wykorzystanie wielu równoległe działających procesorów i I/O podczas operacji agregacji
- ▶ Integracja z motorem SAS SQL.
- ▶ Wielowątkowe przetwarzanie warunków WHERE.
- ▶ Optymalizator warunków WHERE ocenia strategię wybrania podzbioru danych
- ▶ Wsparcie dla operacji “full table scan”, gdy nie ma zdefiniowanych indeksów
- ▶ Uruchamianie wiele wątków, w celu jednoczesnego przetwarzania różnych porcji danych
- ▶ mechanizm WHERE planner costing dla oceny strategii przetwarzania zapytania, w celu wybrania najmniej zasobożernej.

Unikalna technologia indeksów

- ▶ Automatyczne odświeżanie podczas otwierania raportu lub ręcznie podczas Indeks typu Hybrid bitmap łącząca najlepsze cechy indeksu B-tree i bitmapowego.
- ▶ Analiza zawartości tablicy i wybór najlepszego dla niej układ indeksu
- ▶ Budowa i modyfikacja wielu indeksów jednocześnie podczas pojedynczego odczytu tablicy.
- ▶ Utrzymywanie tablicy ze statystykami o indeksach

Optymalizacja zapytań

- ▶ Optymalizacja zapytań typu „correlated”
- ▶ Przejroczysta rekonstrukcja i przekształcenie zagnieżdżonego kodu SQL, w celu stworzenia w locie zbiorów tymczasowych i przyspieszenia złożonych przetwarzań SQL
- ▶ Wykonywanie szybkiego sortowania przy frazach BY
- ▶ Sortowanie podzbiorów danych wybranych na podstawie warunków WHERE lub listy KEEP.
- ▶ Tymczasowy wynik sortowania (sort bin) większy niż 2G, co zmniejsza liczbę takich plików koniecznych do zakończenia sortowania.

Wsparcie dla SQL Pass-Through

- ▶ Przesyłanie zapytania SQL specyficznego dla SPDS-a bezpośrednio do wykonania przez serwer.
- ▶ Uruchamianie równoległe wielu niezależnych zapytań SQL.
- ▶ SQL Pass-Through może być używany dla biblioteki SAS (polecenie LIBNAME).

Zarządzanie i bezpieczeństwo

- ▶ Obsługa dużej liczby użytkowników i rosnących zbiorów danych.
- ▶ Funkcjonalność Automatic aging rollout do ładowania i usuwania danych z bardzo dużych tablic przy użyciu jednostek danych określonych przez stempel czasowy.
- ▶ Wsparcie dla 64-bitów („wide-mode”) pozwalające na alokację i adresowanie powyżej 2G RAM-u za jednym razem
- ▶ Możliwość współpracy z aplikacjami typu desktop pozwalająca im na wykorzystanie mocy obliczeniowej SAS Scalable Performance Data Server.
- ▶ Wyrafinowana implementacja bezpieczeństwa dająca administratorowi kontrolę uprawnień do odczytu, zapisu i modyfikacji tablic na poziomie użytkownika lub grupy

SAS® Scalable Performance Data Server

Wymagania Techniczne

Microsoft Windows (32-bit)

Server OS: Windows NT 4 Server 4.0, Windows 2000, Windows Server 2003

AIX (64-bit)

Release 5.1 or later running on 64-bit Power architecture hardware

HP-UX Itanium (64-bit)

Release 11i (11.11)

Solaris (64-bit)

Sun Solaris 8 or 9

Tru64 UNIX (64-bit)

Version 5.1A or Version 5.1B

Minimum procesorów

Rekomendujemy serwer 4-procesorowy,

Wymagania na RAM

4 GB rekomendowane

Wymagania na przestrzeń dyskową

Dla uzyskania największej wydajności rekomendujemy zastosowanie wielu dysków i kanałów I/O

Opcjonalne oprogramowanie

Base SAS bardzo rekomendowany

Bezpieczeństwo

Poziom użytkownika I grupy

Wspierane protokoły

ODBC



SAS Polska

ul. Gdańska 27/31

01-633 Warszawa

tel. +48 22 560 46 00 do 02

fax. +48 22 560 46 04

www.sas.com/poland