

XIII Konferencja PLOUG
Kościelisko
Październik 2007

Baza danych Oracle11g – nowe cechy

Mariusz Masewicz
Politechnika Poznańska

Mariusz.Masewicz@cs.put.poznan.pl

Abstrakt. Wreszcie doczekaliśmy się informacji na temat tego, co nam przyniesie jedenasta wersja naszej ulubionej bazy danych. Tym razem tradycyjnie ewolucja i rewolucja idą w parze sprawiając, że do ręki otrzymamy produkt, który w wielu miejscach pozytywnie zaskoczy nas swoim działaniem.

Lista nowych cech bazy danych Oracle 11g zawiera prawie 500 pozycji.

Niniejszy referat prezentuje nowe cechy bazy danych Oracle 11g, pokazując jednocześnie propozycje ich zastosowania do rozwiązania typowych problemów, na które napotykają w swojej codziennej pracy administratorzy i programiści bazy danych. Zmiany dotyczą praktycznie wszystkich aspektów pracy z bazą danych - od wykonywania pojedynczych poleceń, poprzez zmianę sposobu patrzenia na składowane dane, do zaawansowanych mechanizmów zarządzania wielkim i wolumenami danych, strojenia bazy danych, czy też zarządzania niezawodnością.

Informacja o autorze: mgr inż. Mariusz Masewicz jest pracownikiem Instytutu Informatyki Politechniki Poznańskiej. Od wielu lat jest też twórcą, administratorem i użytkownikiem wielu aplikacji opartych o bazę danych Oracle. Przez cały ten czas w kręgu jego zainteresowań znajduje się szereg zagadnień związanych z optymalizacją pracy poszczególnych składników architektury Oracle.

Wstęp

Obserwując kierunek, w którym zmierzają zmiany wprowadzane w kolejnych wersjach systemu zarządzania bazą danych Oracle możemy doszukiwać się w nich trzech głównych kierunków:

- wydajna obsługa coraz większych zbiorów danych,
- coraz prostsze administrowanie i strojenie bazy danych,
- coraz silniejszy nacisk na szeroko rozumiane bezpieczeństwo całego systemu.

W najnowszej – jedenastej – wersji bazy danych Oracle lista nowych cech liczy około 500 pozycji. W większości są to rozszerzenia funkcjonalności już istniejących elementów systemu zarządzania bazą danych, pozwalające na efektywniejszą pracę tych elementów. Lista całkiem nowych cech też jest jednak dość pokaźna i obejmuje ona zarówno nowe automaty ułatwiające pracę administratora, nowy sposób działania optymalizatora zapytań, nowe mechanizmy pozwalające zwiększyć efektywność wykonywania zapytań, a także nowe elementy języków SQL, PL/SQL i Java.

W niniejszym artykule postaram się przedstawić zbiór tych nowych cech, które uważam za najbardziej interesujące z punktu widzenia zarówno administratora, jak i programisty bazy danych. W dalszej części artykułu w skrócie omówię pozostałe nowości, które mają sprawić, że nowa wersja bazy danych będzie postrzegana przez jej użytkowników jako produkt jakościowo lepszy zarówno od poprzednich wersji systemów zarządzania bazą danych firmy Oracle, jak i od konkurencyjnych systemów pochodzących od innych dostawców.

1. Pierwsza dziesiątka

Każdy z użytkowników oglądając listę nowych cech danego systemu typuje swoich faworytów. Tak było i w moim przypadku – spośród wszystkich wymienianych nowości wytypowałem dziesięć cech, które chciałem przetestować w pierwszej kolejności. Lista ta zawierała zarówno elementy przydatne administratorowi (Automatic Health Monitoring, Automatic Diagnostic Repository, Hang Manager, zwiększone wsparcie dla szeroko rozumianego bezpieczeństwa bazy danych), ułatwiające życie programiście (rozszerzenia mechanizmów Flashback: Flashback Transaction, Flashback Data Archive, nowe sposoby partycjonowania danych, zwiększone wsparcie dla XML'a, wirtualne kolumny tabel), oraz oczywiście elementy sprawiające, że użytkownik końcowy będzie bardziej zadowolony, gdyż jego zapytania będą wykonywane znacząco szybciej (Result Cache, nowe mechanizmy strojenia zapytań).

Spśród wymienionych wyżej faworytów **pierwsze** miejsce na mojej liście zajęły mechanizmy pozwalające na lepsze strojenie zapytań SQL. Rodzina ta obejmuje między innymi:

- możliwość samodzielnego „strojenia” się zapytań poprzez automatyczne wywoływanie – znanych już z wersji 10 bazy danych – „podpowiadaczy”: SQL Tuning Advisor i SQL Access Advisor;
- możliwość wyliczania statystyk „wielokolumnowych”;
- nowe wskazówki narzędzia SQL Access Advisor – podpowiadające, jak mógłby wyglądać efektywniejszy podział na partycje istniejących już obiektów (tabele, indeksy, perspektywy zmaterializowane);
- nowe podejście do zliczania statystyk – obecnie ponowne zebranie statystyk nie powoduje automatycznego unieważnienia planów wykonania zapytań przechowywanych w obszarze bibliotecznym pamięci serwera;

- historia planów wykonania zapytań, pozwalająca optymalizatorowi porównywać wzrost (lub spadek) wydajności kolejnych wersji planów wykonania tego samego zapytania.

Drugie miejsce na mojej liście przypadło rodzinie narzędzi odpowiadających za monitorowanie „zdrowia” bazy danych, oraz raportowania wszystkich objawów jej „gorszego samopoczucia”. Do tych narzędzi można zliczyć:

- **Automatic Health Monitoring** – narzędzie monitorujące stan poszczególnych bloków danych wykorzystywanych właśnie przez system, stan dziennika powtórzeń, stan segmentów wycofania, oraz stan słownika bazy danych;
- **Automatic Diagnostic Repository** – nowe repozytorium, które zawiera w sobie informacje niezbędne do działania narzędzi diagnostycznych stan bazy danych. Repozytorium to składa się z szeregu plików diagnostycznych, nie będących częścią bazy danych. Częścią ego repozytorium są znane już z wcześniejszych wersji pliki: alert.log (w formacie XML), pliki core, user i background dump;
- **Support Workbench** – element Enterprise Managera, pozwalający na diagnozowanie problemów z bazą danych, wykrytych przez ADR;
- **Incident Packaging Service** – narzędzie pozwalające na zebranie niezbędnych danych opisujących problemy występujące z bazą danych i wysłanie ich do działu pomocy technicznej firmy Oracle;
- **Hang Manager** – rozszerzenie znanego z wersji dziesiątej bazy danych narzędzia Hang Analysis. Pozwala na śledzenie zależności pomiędzy kolejnymi awariami systemu i na tej podstawie wytypowanie prawdopodobnych przyczyn awarii.

Na **trzecim** miejscu umieściłem mechanizmy zachowywania wyników właśnie wykonywanych zapytań: **Result Cache**. Dzięki temu można teraz oznaczyć zapytanie, funkcję w języku PL/SQL lub wywołanie poprzez interfejs OCI jako działanie, którego wynik należy zapamiętać w pamięci SGA. Dzięki temu inne zapytania (funkcje, lub wywołania OCI) mogą korzystać z tych zbuforowanych wyników jak z dostępnego prawie natychmiast źródła danych (np. traktowanych jako dane cząstkowe do sporządzenia większego agregatu, czy raportu).

Kolejne – **czwarte** – miejsce zajmują nowe narzędzia ułatwiające proces strojenia bazy danych:

- **Database Replay** – to narzędzie, które pozwala zapamiętać zbiór zapytań wydawanych do bazy danych, a następnie po dokonaniu zmian w bazie danych pozwala na powtórzenie tych zapytań, aby zweryfikować pozytywny (lub negatywny) wpływ wprowadzonych zmian na wydajność bazy danych;
- **SQL Performance Analyzer** – uzupełniające działanie Database Replay, umożliwiając „zmierzenie” wydajności wykonania zbioru zapytań (na przykład zbioru wykorzystywanego przez Database Replay do testowania wprowadzonych zmian w bazie danych).

Na **piątym** miejscu znalazły się mechanizmy zapewniające bazie danych bezpieczniejsze i łatwiejsze zarządzanie użytkownikami. Po raz pierwszy w historii większość mechanizmów zwiększających to bezpieczeństwo jest automatycznie włączone. Do tych mechanizmów należą:

- automatycznie włączona funkcja weryfikująca stopień trudności haseł użytkowników, hasła są wrażliwe na wielkość wprowadzanych liter;
- domyślny profil (DEFAULT) posiada zdefiniowane ustawienia dla parametrów: password grace time, life time, and lock time, failed login attempts;

- domyślnie jest włączony mechanizm auditu pewnych działań wykonywanych z uprawnieniami DBA (CREATE ANY PROCEDURE, GRANT ANY PRIVILEGE, DROP USER). Wyniki tego auditu są domyślnie przechowywane wewnątrz bazy danych;
- Enterprise Manager został wyposażony w funkcjonalność ułatwiającą zarządzaniem mechanizmami: External Security Module (ESM), Fine-Grained Auditing (FGA), oraz Row-Level Security (RLS) policies.

Miejsce **szóste** należy do kolejnych rozszerzeń systemu zapewniania bezpieczeństwa bazy danych. Tym razem jest to mechanizm **SecureFiles** – pozwalający na bezpieczne przechowywanie w bazie danych „dużych” obiektów (obrazy, teksty, filmy, XML). Mechanizm ten zapewnia bezpieczne składowanie tych obiektów, a także dodatkowe funkcje, takie jak wykrywanie i usuwanie duplikatów, szyfrowanie danych. Dodatkowo w Oracle 11g dane mogą być szyfrowane na poziomie całej przestrzeni tabel.

Siódme miejsce to rozszerzone mechanizmy partycjonowania danych. Począwszy od wspomnianego już wcześniej rozbudowania funkcjonalności SQL Access Advisor, poprzez nowe metody partycjonowania wielopoziomowego: Range-Range, List-Range, List-Hash, oraz List-List. Kolejne rozszerzenia mechanizmu partycjonowania to:

- partycjonowanie interwałowe – czyli mechanizm sprawiający, że jeżeli kluczem partycjonowania będzie okres czasu (tydzień, miesiąc, kwartał, rok) lub zakres liczb (DATE lub NUMBER) – to w chwili pojawiania się danych z nowego okresu system automatycznie utworzy nową partycję;
- partycjonowanie bazujące na wirtualnych kolumnach – począwszy od wersji 11g tabela może posiadać „kolumny wirtualne” (czyli kolumny, których wartość nie jest wprowadzana wprost do tabeli, a jest zdefiniowana jako wyrażenie wyliczane na podstawie wartości innych kolumn tej tabeli – na przykład kolumna CENA_BRUTTO zdefiniowana jako wartość kolumny CENA_NETTO pomnożona przez wartość podatku VAT. Wartość kolumny wirtualnej jest składowana w bazie danych i automatycznie pielęgnowana przez system). Kluczem partycjonowania tabeli może być kolumna wirtualna;
- partycjonowanie bazujące na referencjach – mechanizm przydatny wówczas, kiedy tabele są powiązane więzami referencyjnymi i tabela nadrzędna jest partycjonowana według innego klucza niż ograniczenie referencyjne łączące te tabele. Wówczas partycje tabeli podrzędnej zawierają tylko te wiersze, które odpowiadają wierszą w analogicznej partycji założonej dla tabeli nadrzędnej;
- przenoszalne partycje – mechanizm pozwalający na przenoszenie pomiędzy bazami danych pojedynczych partycji będących częścią tabeli partycjonowanej.

Na kolejnym – **ósmym** – miejscu uplasowały się rozszerzenia mechanizmu ASM (Automatic Storage Management).

- mechanizm **ASM Rolling Upgrades** – pozwalający na aktualizowanie kolejnych instancji klastra aplikacyjnego (RAC) minimalizując czas niedostępności aktualizowanego serwisu;
- **Fast Mirror Resynchronization** – mechanizm, który w przypadku awarii dysku (rozłączenia równoczesnego zapisu do dysków) w środowisku z utrzymywaniem kilku jego lustrzanych kopii pozwala na szybkie odtwarzanie tylko tych ekstentów, które były modyfikowane na sprawnych dyskach;
- **Preferred Mirror Read** – mechanizm pozwalający na wskazanie tego dysku (w architekturze z wieloma dyskami powiązanymi w zbiory lustrzanych kopii), z którego będą odczytywane dane. Rozwiązanie przydatne w zastosowaniach typu klastrer aplikacyjny.

cyjny, gdzie można wskazać lokalizacje najbliższej/najszybszej macierzy dyskowej przechowującej jedną z kopii potrzebnych danych;

- możliwość zmiany jednostki alokacji dla obiektów zarządzanych przez ASM;
- zwiększona funkcjonalność narzędzia ASMCMD.

Miejsce **dziewiąte** to zmiany mające zwiększyć wydajność wykonywania kodu w języku PL/SQL:

- implementacja obsługi standardu XA – możliwość obsługi transakcji wykonywanych z wykorzystaniem wielu zarządców zasobów (na przykład różne systemy zarządzania bazą danych);
- poprawiona obsługa sekwencji – od tej pory sekwencje mogą być traktowane jako zmienne:

```
DECLARE
    n NUMBER := Seq.Nextval;
BEGIN
```

- rozszerzenie rozmiaru kodu dopuszczalnego dla dynamicznego SQLa poza granicę dotychczasowych 32K. Dzięki temu możliwe jest obsługiwane przez mechanizmy dynamicznego SQLa danych typu LOB;
- polecenie CONTINUE – które podobnie jak w innych językach programowania – pozwala na zakończenie wykonywania bieżącego obiegu pętli i rozpoczęcie wykonywania kolejnego;
- szybkie wyzwalacze i inne elementy języka PL/SQL – według zapewnień producenta kod PL/SQL ma być wykonywany średnio o 25% szybciej niż analogiczny kod uruchomiony pod kontrolą starszej bazy danych;
- typ danych SIMPLE_INTEGER – typ zgodny z typem danych INTEGER obsługiwany bezpośrednio przez procesory wszystkich komputerów – dzięki niemu programy kompilowane natywnie działają wydajniej;
- wprowadzenie kontroli sekwencji uruchamiania wyzwalaczy bazy danych;
- natywna kompilacja (także dla procedur w języku Java) bez konieczności używania zewnętrznych kompilatorów;
- nowe funkcje do obsługi wyrażeń regularnych: REGEXP_COUNT.
- możliwość stosowania różnych notacji parametrów wywołania funkcji – także dla języka SQL: `SELECT f(1, pn=>3) FROM dual`

Rozszerzone mechanizmy archiwizacji danych zajęły miejsce **dziesiąte**. Mechanizmy te sprawiają, że kopie bezpieczeństwa bazy danych (wykonywane przy pomocy różnych ustawień programów do archiwizacji – RMAN, DPDUMP) mogą być w całości, bądź też w zdefiniowanej części szyfrowane i kompresowane. Spójna kopia bazy danych nie zawiera już w sobie obszarów UNDO.

2. Dalsze miejsca – czyli kolejne ciekawe nowości

W kolejnej części artykułu wymienię te cechy bazy danych Oracle 11g, których wprowadzenie nie wywołało u mnie tak entuzjastycznego przyjęcia, jak poprzednie, ale również one stanowią swego rodzaju przełom w spojrzeniu na nową bazę danych.

- Automatic Memory Tuning – strojenie przydziału pamięci operacyjnej na obszary PGA i SGA od zawsze stanowiło wyzwanie dla administratora. Wcześniejsze wersje bazy danych pozwalały w mniejszym (Oracle 9i) lub większym (Oracle 10g) stopniu zautomatyzować proces trojenia obszaru PGA lub też SGA. W jedenastej wersji bazy danych Oracle – administrator może ten proces całkowicie zautomatyzować, ustawiając tylko jeden parametr konfiguracyjny pozwalający określić ilość pamięci dla instancji i pozostałych procesów;
- rozszerzone mechanizmy kompresji tabel i indeksów – dzięki nowym algorytmom mają działać wydajniej nawet w środowiskach typu OLTP;
- wydajna obsługa sieciowych urządzeń składających dane (macierzy NAS) – **Direct NFS Klient**;
- rozszerzone wsparcie dla języka XML – same dokumenty dalej mogą być przechowywane w bazie danych jako dane typu CLOB – ale za to dodano obsługę zapytań XQuery oraz SQL XML, binarne wsparcie dla dokumentów XML dostępny z poziomu języka Java;
- operacja PIVOT i UNPIVOT w języku SQL – pozwalające na przeglądanie i prezentację danych wielowymiarowych w odpowiednim przekroju;
- rozszerzone mechanizmy informowania o błędach i ostrzeżeniach w języku PL/SQL;
- dalsze usprawnianie mechanizmu perspektyw zmaterializowanych i bazującego na nich mechanizmu przepisywania za pytań, które obecnie działają poprawnie także dla sytuacji ze zdalnymi tabelami, czy też perspektywami *inline*;
- nowy, wydajniejszy algorytm sortowania danych;
- wsparcie dla standardu UNICODE 5.0;
- pełne wsparcie dla standardu Java 5.0 (kompilator w wersji 1.5);
- sterownik JDBC rozszerzony dla spełniania wymagań standardu 4.0.
- sterownik JDBC wspiera nowe typy danych: `java.sql.XMLType`, `java.sql.RowID`, `java.sql.NClob`, `ANYTYPE`, `ANYDATA`, `ANYDATASET`;
- coraz większe wsparcie dla platformy .NET (i narzędzia Visual Studio 2005) – trzy tryby debugowania aplikacji: Direct Database Debugging, Multi-Tier Application Debugging, External Application Debugging;
- nowy sterownik dla języka PHP;
- rozszerzenie mechanizmu Flashback o dwie nowe opcje: **Flashback Transaction** – czyli możliwość wycofanie jednej lub wielu transakcji bazując na danych z segmentów wycofania; i **Flashback Data Archive** – pozwala zdefiniować sposób przechowywania danych w segmentach wycofania i prezentuje te dane poprzez szereg perspektyw na przykład do celów auditu bezpieczeństwa bazy danych;
- rozszerzenie funkcjonalności APEX w wersji 3.0: Wsparcie dla technologii AJAX, możliwość przesuwania i upuszczania elementów, wykresy w technologii Flash, przyjazne adresy URL, narzędzia do migracji aplikacji z technologii Microsoft Access, buforowanie wygenerowanych stron WWW, drukowanie dokumentów PDF,

Podsumowanie

Powyższy przegląd cech nowych bazy danych Oracle 11g powinien pokazywać, że jest to produkt, który z pewnością ułatwi życie swoim użytkownikom. Wymienione tu nowe cechy bazy danych powinny być przydatne wszystkim grupom użytkowników – począwszy od administratorów, poprzez programistów, a na użytkownikach końcowych przestając. Można powiedzieć, że Oracle 11g „upraszcza” rozwój infrastruktury, jednocześnie lepiej zarządzając jej zasobami.

System zarządzania bazą danych Oracle 11g umożliwi natychmiastową reakcję na problemy bazodanowe bez utraty stabilności.

Środowisko bazodanowe w Oracle 11g jest bardziej skalowalne, należycie zabezpieczone, i wysoko dostępne.

Wszystkie te cechy wymagają oczywiście dokładnego przetestowania, zapoznania się z ich wymaganiami i możliwościami aby na końcu móc pozytywnie odpowiedzieć sobie na pytanie: czy już mam dość wiedzy o Oracle 11g aby uruchomić ten system na środowisku produkcyjnym?

Bibliografia

http://download.oracle.com/docs/cd/B28359_01/server.111/b28279/toc.htm - Oracle® Database New Features Guide 11g Release 1 (11.1)

http://download.oracle.com/docs/cd/B28359_01/server.111/b28310/toc.htm - Oracle® Database Administrator's Guide 11g Release 1 (11.1)

http://download.oracle.com/docs/cd/B28359_01/server.111/b28286/toc.htm - Oracle® Database SQL Language Reference 11g Release 1 (11.1)