

Ćwiczenie 3.

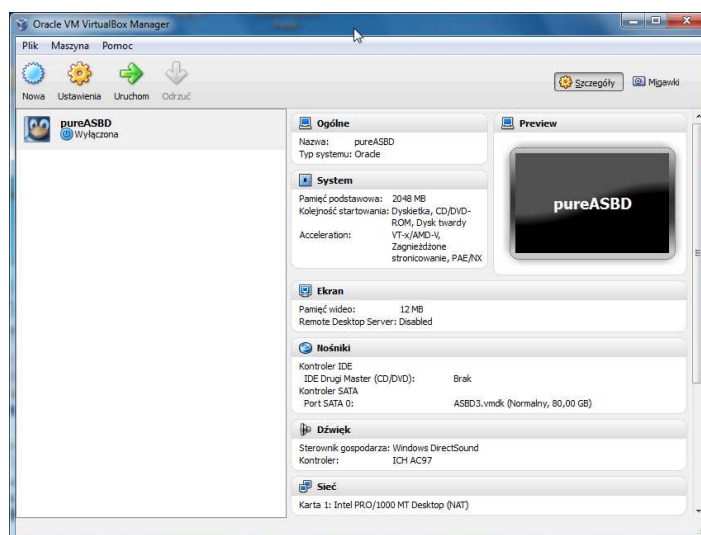
Funkcje analityczne

1. Uruchomienie i skonfigurowanie środowiska do ćwiczeń

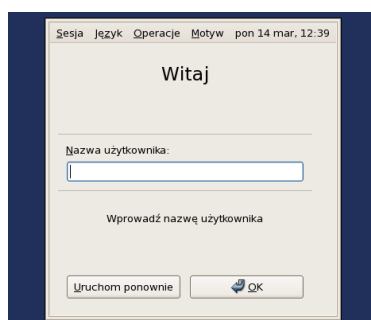
Czas trwania: 15 minut

Zadaniem niniejszych ćwiczeń jest przedstawienie podstawowych zagadnień dotyczących wykorzystywania funkcji analitycznych w języku SQL.

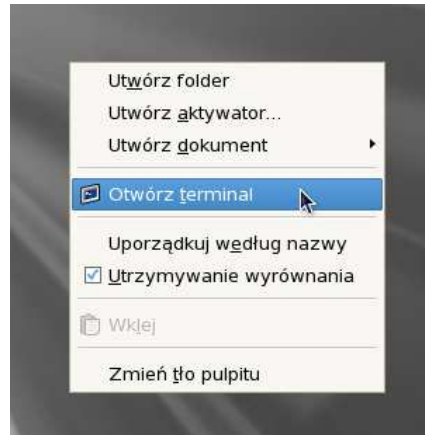
1. Uruchom środowisko wirtualizacji – kliknij na umieszczoną na pulpicie komputera-gospodarza ikonę *Oracle VM VirtualBox*.



2. Spróbuj uruchomić maszynę wirtualną. W tym celu zaznacz w lewym panelu środowiska pozycję **ASBD** i naciśnij umieszczony na pasku narzędzi przycisk **Uruchom** (możesz również wybrać pozycję **Uruchom** z menu kontekstowego, dostępnego po kliknięciu prawym klawiszem myszy na pozycji **ASBD**).
3. Po pomyślnym uruchomieniu maszyny wirtualnej powinien zostać wyświetlony ekran logowania (jeśli ekran maszyny wirtualnej nie zajmuje całego ekranu komputera-gospodarza, użyj odpowiedniego skrótu klawiszowego aby to zmienić).



4. Zaloguj się do systemu operacyjnego maszyny wirtualnej jako użytkownik *Oracle* z hasłem *oracle*. Podaj powyższe informacje i naciśnij przycisk **OK**.
5. Uruchom terminal. Można to wykonać przez kliknięcie prawym klawiszem myszki na pulpit i wybranie z menu kontekstowego pozycji **Otwórz terminal**.



6. Sprawdź wartości zmiennych środowiskowych dotyczących instancji *Oracle*. Wykorzystaj w tym celu polecenie: `set | grep ORACLE`

```
oracle@localhost:~  
Plik Edycja Widok Terminal Zakładki Pomoc  
[oracle@localhost ~]$ set | grep ORACLE  
ORACLE_BASE=/u01/app/oracle  
ORACLE_HOME=/u01/app/oracle/product/11.2.0/dbhome_1  
ORACLE_SID=baza01  
[oracle@localhost ~]$
```

7. Jeśli zmienne środowiskowe nie są ustawione poprawnie, to wczytaj ich wartości z pliku `.bashrc` umieszczonym w Twoim katalogu domowym

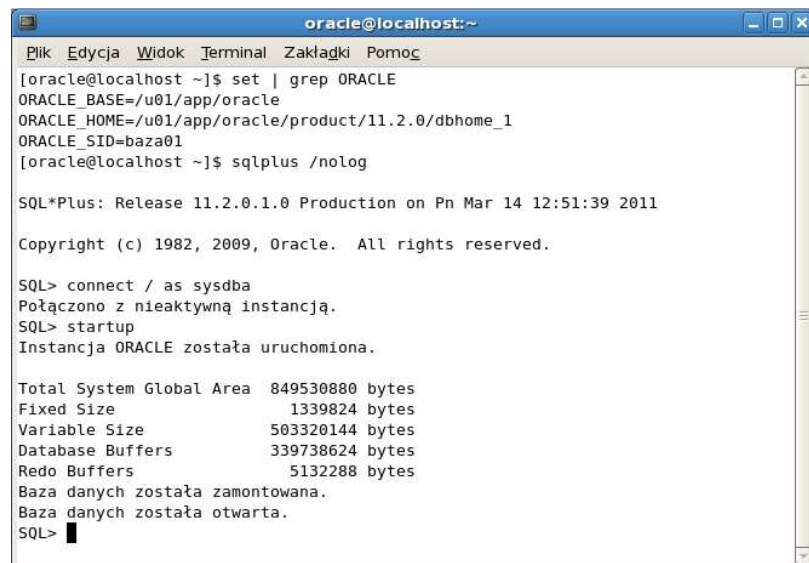
```
[oracle@localhost ~] source .bashrc
```

8. Jeśli zmienne środowiskowe są ustawione poprawnie, uruchom program *sqlplus*. Wykorzystaj opcję *nolog*: `sqlplus /nolog`

```
oracle@localhost:~  
Plik Edycja Widok Terminal Zakładki Pomoc  
[oracle@localhost ~]$ set | grep ORACLE  
ORACLE_BASE=/u01/app/oracle  
ORACLE_HOME=/u01/app/oracle/product/11.2.0/dbhome_1  
ORACLE_SID=baza01  
[oracle@localhost ~]$ sqlplus /nolog  
  
SQL*Plus: Release 11.2.0.1.0 Production on Pn Mar 14 12:51:39 2011  
Copyright (c) 1982, 2009, Oracle. All rights reserved.  
  
SQL>
```

9. Następnie zaloguj się korzystając z autoryzacji użytkownika administracyjnego przez system operacyjny. Wykonaj w tym celu polecenie: `connect / as sysdba`

Następnie uruchom bazę danych poleceniem **startup**



```
oracle@localhost:~  
Plik Edycja Widok Terminal Zakładki Pomoc  
[oracle@localhost ~]$ set | grep ORACLE  
ORACLE_BASE=/u01/app/oracle  
ORACLE_HOME=/u01/app/oracle/product/11.2.0/dbhome_1  
ORACLE_SID=baza01  
[oracle@localhost ~]$ sqlplus /nolog  
  
SQL*Plus: Release 11.2.0.1.0 Production on Pn Mar 14 12:51:39 2011  
  
Copyright (c) 1982, 2009, Oracle. All rights reserved.  
  
SQL> connect / as sysdba  
Połączono z nieaktywną instancją.  
SQL> startup  
Instancja ORACLE została uruchomiona.  
  
Total System Global Area 849530880 bytes  
Fixed Size 1339824 bytes  
Variable Size 503320144 bytes  
Database Buffers 339738624 bytes  
Redo Buffers 5132288 bytes  
Baza danych została zamontowana.  
Baza danych została otwarta.  
SQL>
```

10. Pobierz ze strony WWW studium podyplomowego plik `spbd_funkcje.dmp`
11. Uruchom drugi terminal i zaimportuj całą zawartość pobranego pliku do schematu użytkownika `scott`. Możesz się posłużyć poniższym poleceniem.

```
imp file=spbd_funkcje.dmp userid=scott/tiger full=y
```

12. Zaloguj się jako użytkownik `scott` z hasłem `tiger`. Odczytaj ze słownika bazy danych nazwy tabel umieszczonych w schemacie. Zapoznaj się ze strukturą tabeli zawierającej przykładowe dane o transakcjach finansowych oraz wyświetl zawartość tabeli.

```
SELECT table_name FROM user_tables;
```

```
DESCRIBE transakcje
```

```
SELECT * FROM transakcje;
```

2. Funkcje analityczne

Czas trwania: 60 minut

Przystąpimy teraz do przeglądu funkcji analitycznych oferowanych przez RDBMS Oracle

1. Wyświetl ranking (rzadki i gęsty) kwot transakcji dla konta '11-11111111' z podziałem na kategorie operacji.

```
SELECT RANK() OVER (PARTITION BY kategoria
                    ORDER BY kwota DESC) RANK,
       DENSE_RANK() OVER (PARTITION BY kategoria
                          ORDER BY kwota DESC) dense_rank,
       kategoria, data, kwota
FROM   transakcje
WHERE  nr_konta = '11-11111111'
ORDER BY kategoria, kwota DESC;
```

2. Wyświetl percentyle oraz pozycje procentowe rankingu wpłat pensji na konto '11-11111111'

```
SELECT CUME_DIST() OVER (ORDER BY kwota) cume_dist,
       PERCENT_RANK() OVER (ORDER BY kwota) percent_rank, kwota
FROM   transakcje
WHERE  nr_konta = '11-11111111'
AND    kategoria = 'PENSJA'
ORDER BY kwota;
```

3. Wyświetl podział wpłat i wypłat związanych z pensją i rachunkiem za telefon na cztery grupy w zależności od wysokości wpłaty/wypłaty.

```
SELECT NTILE(4) OVER (ORDER BY kwota) ntile,
       ROW_NUMBER() OVER (ORDER BY kwota) row_number,
       kwota, data, kategoria
FROM   transakcje
WHERE  nr_konta = '11-11111111'
AND    kategoria IN ('PENSJA', 'RACHUNEK ZA TELEFON')
ORDER BY kwota;
```

4. Znajdź trzy najwyższe wpłaty

```
SELECT kwota, data, nr_konta
FROM (
    SELECT kwota, data, nr_konta,
           RANK() OVER (ORDER BY kwota DESC) AS ranking
    FROM   transakcje)
WHERE ranking <= 3;
```

5. Dla każdej transakcji przedstaw jej datę, kwotę na jaką opiewała oraz saldo po wykonaniu operacji, średnią kwotę operacji z ostatniego roku, minimalną kwotę z 3 ostatnich operacji, oraz liczbę operacji wykonanych od 6 miesięcy wstecz do 6 miesięcy po transakcji. Zapytanie ogranicz do konta nr '11-11111111'

```
SELECT kwota, data,
       SUM(kwota) OVER (ORDER BY data) saldo,
       AVG(kwota) OVER (ORDER BY data RANGE INTERVAL '12' month
PRECEDING) avg12,
       MIN(kwota) OVER (ORDER BY data rows 3 PRECEDING) min3,
       COUNT(*) OVER (ORDER BY data RANGE BETWEEN interval '6'
month PRECEDING AND interval '6' month FOLLOWING) count6_6
FROM transakcje
WHERE nr_konta = '11-11111111'
ORDER BY data;
```

6. Dla każdej transakcji przedstaw jej kwotę, datę i kategorię, oraz średnią kwotę operacji wchodzących w skład tej samej kategorii i udział kwoty transakcji do wszystkich transakcji z tej samej kategorii. Zapytanie ogranicz do konta nr '11-11111111'

```
SELECT kwota, data, kategoria,
       AVG(kwota) OVER (PARTITION BY kategoria) AS avg_k,
       kwota/SUM(kwota) OVER
(PARTITION BY kategoria) AS ratio_to_report
FROM transakcje
WHERE nr_konta = '11-11111111'
ORDER BY kategoria, data;
```

7. Wyświetl salda kroczące dla konta nr '11-11111111'. Wykorzystaj funkcję LEAD do znalezienia daty następnej operacji na koncie (w wyniku której zmieni się saldo)

```
SELECT kwota,
       SUM(kwota) OVER (ORDER BY data) AS saldo,
       data AS od_dnia,
       LEAD(data,1) OVER (ORDER BY data) AS do_dnia
FROM transakcje
WHERE nr_konta = '11-11111111'
ORDER BY data;
```

8. Dla każdego roku wyświetl kwotę i datę największego przychodu.

```
SELECT TO_CHAR(data,'yyyy'),
       MAX(kwota) przychod,
       MAX(data) keep (DENSE_RANK LAST ORDER BY kwota) dzien
FROM transakcje
WHERE nr_konta = '11-11111111'
GROUP BY TO_CHAR(data,'yyyy');
```

9. Dla każdego roku wyświetl medianę wartości kwotowej transakcji na koncie nr '11-11111111'

```
SELECT TO_CHAR(data,'yyyy') rok,  
       PERCENTILE_DISC(0.5) WITHIN GROUP (ORDER BY kwota) disc,  
       PERCENTILE_CONT(0.5) WITHIN GROUP (ORDER BY kwota) cont  
FROM   transakcje  
WHERE  nr_konta = '11-11111111'  
GROUP BY TO_CHAR(data,'yyyy');
```

10. Sprawdź, w którym miejscu w rankingu znalazłaby się wpłata 3000 zł.

```
SELECT RANK(3000) WITHIN GROUP  
       (ORDER BY SUM(kwota) DESC) GDZIE_3000  
FROM   transakcje  
GROUP BY nr_konta, TO_CHAR(data,'YYYY');
```

11. Porównaj efektywność znalezienia trzech największych transakcji za pomocą funkcji analitycznej i za pomocą tradycyjnego podzapytania SQL

```
SET AUTOTRACE ON EXPLAIN
```

```
SELECT kwota, data, nr_konta  
FROM ( SELECT kwota, data, nr_konta,  
             RANK() OVER (ORDER BY kwota DESC) r  
       FROM   transakcje)  
WHERE r <= 3;
```

```
SELECT kwota, data, nr_konta  
FROM   transakcje T1  
WHERE 3 >= ( SELECT COUNT(*)  
            FROM   transakcje  
            WHERE kwota >= T1.kwota);
```

3. Funkcje analityczne: ćwiczenia samodzielne

Funkcje rankingu

1. Dla każdego konta wyświetl saldo na rachunku i miejsce w rankingu kont ustalonego w oparciu o wielkość salda

NR_KONTA	SALDO	RANKING
-----	-----	-----
22-22222222	9499,5	1
11-11111111	7479,5	2
33-33333333	6869,5	3

2. Dla konta 11-11111111 podziel wszystkie transakcje na cztery równe części pod względem czasu wykonania. Czy wyniki są niedeterministyczne?

NTILE	KWOTA	DATA	KATEGORIA
-----	-----	-----	-----
1	1500	97/01/03	PENSJA
1	-100	97/01/12	WYPŁATA W BANKOMACIE
1	750	97/02/02	UMOWA O DZIEŁO
1	-150	98/05/22	RACHUNEK ZA TELEFON
2	1800	99/11/04	PENSJA
2	1800	99/12/02	PENSJA
2	-300	99/12/24	WYPŁATA W BANKOMACIE
2	-120,5	99/12/24	RACHUNEK ZA PRĄD
3	-150	00/01/19	RACHUNEK ZA TELEFON
3	900	00/05/03	UMOWA O DZIEŁO
3	-600	01/07/16	RACHUNEK ZA PRĄD
3	-150	01/07/24	WYPŁATA W BANKOMACIE
4	1400	01/09/03	PENSJA
4	-250	01/09/29	RACHUNEK ZA TELEFON
4	1650	01/12/03	PENSJA
4	-500	02/01/05	WYPŁATA W BANKOMACIE

3. Dla każdego z kont znajdź najwcześniej wykonaną transakcję.

NR_KONTA	KWOTA	DATA	RANKING
-----	-----	-----	-----
11-11111111	1500	97/01/03	1
22-22222222	1000	97/01/12	1
33-33333333	1350	95/10/03	1

4. Dla każdego z kont znajdź rok, w którym wykonano największą liczbę transakcji

NR_KONTA	ROK	RANKING
-----	-----	-----
11-11111111	2001	1
22-22222222	2001	1
33-33333333	2000	1

Funkcje okna

5. Dla każdej transakcji przedstaw datę, kwotę na jaką opiewała, oraz średnią kwotę z transakcji bankowych mających miejsce co najwyżej 6 miesięcy wcześniej. Czy był taki moment, w którym średnia ta była ujemna?

DATA	KWOTA	SREDNIA
-----	-----	-----
95/10/03	1350	1350
96/07/02	670	670
97/01/03	1500	1500
97/01/12	1000	1250
97/01/12	-100	800
97/02/02	750	787,5
97/02/12	-90	612
97/09/02	830	830
. . .		

6. Dla każdej transakcji przedstaw datę, kwotę na jaką opiewała i stan bankowego skarbca w momencie jej zaksięgowania. Czy kiedykolwiek stan ten był ujemny? Jaki był stan na koniec dnia 99/12/24?

DATA	KWOTA	STAN
-----	-----	-----
95/10/03	1350	1350
96/07/02	670	2020
97/01/03	1500	3520
97/01/12	1000	4520
97/01/12	-100	4420
97/02/02	750	5170
97/02/12	-90	5080
97/09/02	830	5910
98/03/22	-170	5740
98/04/03	1650	7390
. . .		

7. Dla każdego konta znajdź transakcje, które doprowadziły do stanu debetowego. Podaj numer konta, datę transakcji, kwotę na jaką opiewała oraz saldo rachunku po transakcji.

nie wybrano żadnych wierszy

Funkcje raportujące

8. Dla każdego konta wyświetl daty i kwoty wynikające z transakcji o kategorii PENSJA. Dla każdej transakcji podaj kwotę, o jaką różni się kwota transakcji od średniej pensji zaksięgowanej na tym koncie.

NR_KONTA	DATA	KWOTA	ROZNICA
11-11111111	97/01/03	1500	-130
11-11111111	99/11/04	1800	170
11-11111111	99/12/02	1800	170
11-11111111	01/09/03	1400	-230
11-11111111	01/12/03	1650	20
22-22222222	98/04/03	1650	-30
22-22222222	97/01/12	1000	-680
22-22222222	01/05/04	1980	300
22-22222222	01/07/02	2090	410
22-22222222	02/11/03	1540	-140
22-22222222	03/04/03	1820	140
33-33333333	95/10/03	1350	-134
. . .			

Funkcje LAG/LEAD

9. Dla każdego roku podaj rok, sumę kwot zaksięgowanych w danym roku, oraz różnicę pomiędzy sumą kwot zaksięgowanych w roku bieżącym a sumą kwot zaksięgowanych w roku wcześniejszym.

ROK	ROK_BIEZACY	ROZNICA
1995	1350	
1996	670	-680
1997	3890	3220
1998	4530	640
1999	3419	-1111
2000	3019,5	-399,5
2001	5100	2080,5
2002	50	-5050
2003	1820	1770

Funkcje rankingu hipotetycznego

10. Gdyby w jakimś roku suma kwot zaksięgowanych osiągnęła 3000, to który z kolei byłby to rok w rankingu lat pod względem sumy zaksięgowanych kwot?

```
      KTORY
-----
      6
```

Funkcje analityczne a wydajność

11. Czy potrafiłbyś napisać zapytanie będące odpowiednikiem zadania pierwszego bez użycia funkcji analitycznych? Jeśli tak, to porównaj liczbę operacji I/O (consistent gets) konieczną do wykonania obu wariantów rozwiązań dla tego samego zadania.

NR_KONTA	SALDO	RANKING
-----	-----	-----
11-11111111	7479,5	2
22-22222222	9499,5	1
33-33333333	6869,5	3