

Mascarea datelor în Oracle

Cuvinte cheie: <ul style="list-style-type: none">• Mascare date	<ul style="list-style-type: none">• Oracle Data Pump• Export & import de date
---	--

1. Introducere

- Atunci când sunt instalate în sistemul de producție, aplicațiile se execută pe date reale. Însă, în perioada de dezvoltare și testare a sistemului, programatorii au nevoie de o bază de date de test.
- Adesea datele de test sunt extrase din baze de date din producție pentru a simula un cadru de testare cât mai apropiat de cel real de exploatare.
- Există date sensibile la politicile de confidențialitate care nu pot fi date pur și simplu pentru testare. Aceste date includ: CNP-uri, PIN-uri de card, locații GPS, codul angajatului etc.

Se pune problema mascării datelor confidențiale.

- Există mai multe instrumente furnizate de Oracle (*Oracle Data Masking Pack*, *Oracle Data Pump*) și există, de asemenea, posibilitatea implementării unor subprograme adaptate la nevoile aplicației.
- Indiferent de metoda aleasă, există câteva aspecte importante ce necesită o atenție deosebită:
 - gradul în care metoda aleasă păstrează distribuția valorilor datelor. Specialiștii atrag atenția asupra faptului ca performanțele aplicației se pot modifica radical în cazul în care prin mascare se modifică histograma datelor dintr-o coloană și/sau lungimea valorilor dintr-o coloană;
 - păstrarea integrității referențiale: în cazul în care trebuie mascată o cheie primară, trebuie să se asigure modificarea corespunzătoare, în cascadă, a cheilor externe ce o referă;
 - efortul necesar operației de mascare a datelor: în cazul permutării diferențiate a valorilor pe coloane diferite, trebuie realizat un număr de operații *full scan* egal cu numărul coloanelor din tabel.

2. Oracle Data Pump

- În continuare se va prezenta metoda de mascare a datelor în Oracle folosind **Oracle Data Pump** cu exemplificare pe aplicația de *e-learning* discutată în cadrul laboratoarelor precedente.
- Sub Windows, regăsim fișierele executabile *expdp.exe* și *impdp.exe* în folderul BIN din

calea de instalare a sistemului Oracle (de exemplu, C:\app2\<user>\product\11.2.0\dbhome_1\bin sau D:\app\<user>\virtual\product\11.2.0\dbhome_1\bin). Alternativ, se pot crea *job*-uri folosind API specific (<https://oracle-base.com/articles/10g/oracle-data-pump-10g#GettingStarted>).

- Pentru mascarea datelor cu ajutorul celor două executabile menționate mai sus, vom parcurge următorii pași:

1) Crearea unui pachet PL/SQL în care includem funcții de mascare a datelor

2) Export al tabelelor dorite aplicând funcția de mascare din pachet:

```
expdp user/parola tables=tabel1,tabel2,...
remap_data=tabel1.col_x:pachet.functie_de_mascare
remap_data=tabel2.col_y:pachet.functie_de_mascare2 ...
directory=NUME_DIR dumpfile=NUME_FIS.dmp
```

3) Import al tabelelor cu date mascate:

```
impdp user/parola directory= NUMEDIR dumpfile=NUMEFIS.dmp
TABLES= tabela1,tabela2,... remap_table= tabela1:alias1
remap_table=tabela2:alias2 ...
```

Exercițiu:

Se consideră tabelele STUDENT și REZOLVA având următoarea structură:

SQL> desc student;		
Name	Null?	Type
ID	NOT NULL	NUMBER
NUME		VARCHAR2(30)
PRENUME		VARCHAR2(30)
ANUL		NUMBER
SPECIALIZARE		VARCHAR2(3)
GRUPA		NUMBER

SQL> desc rezolva		
Name	Null?	Type
COD_STUDENT	NOT NULL	NUMBER
COD_TEMA	NOT NULL	NUMBER
DATA_UPLOAD		DATE
NOTA		NUMBER(4,2)
DATA_CORECTARE		DATE

Să se utilizeze mascarea datelor pentru coloanele *id*, *nume*, *prenume*, *grupa* din tabela STUDENT. Se cunoaște următorul fapt: *cod_student* din tabela REZOLVA este cheie externă și referă cheia primară (*id*) din tabela STUDENT.

Observație: Tabelele sunt populate cu următoarele date inițiale:

```
insert into student values(135,'Avramescu','Anton',5,'Inf',531);
insert into student values(212,'Antim','Tudor',5,'Inf',532);
insert into student values(314,'Tinca','Ana',5,'Inf',531);
insert into student values(411,'Caludescu','Aristida',5,'Inf',532);
```

```
insert into materie values(1,'SecBD');
insert into tema values(1,1,SYSDATE-45,1);
insert into tema values(2,1,SYSDATE-30,1);
insert into tema values(3,1,SYSDATE+7,2);
insert into tema values(4,1,SYSDATE+28,1);
```

```
insert into rezolva values(135,1,sysdate-50,null,null);
```

```

insert into rezolva values(212,1,sysdate-45,null,null);
insert into rezolva values(135,2,sysdate-35,null,null);
insert into rezolva values(212,2,sysdate-35,null,null);
insert into rezolva values(314,2,sysdate-30,null,null);
insert into rezolva values(135,3,sysdate,null,null);

```

```
COMMIT;
```

```
SQL> select * from student;
```

ID	NUME	PRENUME	ANUL	SPE	GRUPA
135	Avramescu	Anton	5	Inf	531
212	Antim	Iudor	5	Inf	532
314	Tinca	Ana	5	Inf	531
411	Caludescu	Aristida	5	Inf	532

```
SQL> select * from rezolva;
```

COD_STUDENT	COD_TEMA	DATA_UPLO	NOTA	DATA_CORE
135	1	04-NOV-12		
212	1	09-NOV-12		
135	2	19-NOV-12		
212	2	19-NOV-12		
314	2	24-NOV-12		
135	3	24-DEC-12		

```
6 rows selected.
```

Rezolvare:

- **Pasul 1**) Crearea unui pachet PL/SQL în care includem funcții de mascare a datelor
 - Se dorește păstrarea primului caracter din *nume* și din *prenume*, iar pentru *id* și *grupa* se dorește păstrarea primei cifre.
 - Vom folosi o funcție supraîncărcată în pachet pentru șiruri, respectiv pentru cheia primară *id*, și o altă funcție pentru mascarea grupei.
 - Întrucât *id* este cheie primară referită, vom folosi un tablou indexat care să rețină pe poziția *id* valoarea mascată. Același element din tablou va fi folosit și pentru mascarea lui *cod_student* din REZOLVA.

```

create or replace package pachet_mascare is
    function f_mascare(sir varchar2) return varchar2;
    function f_mascare(nr number) return number;
    function f_mascaregrupa(nr number) return number;
end;
/

create or replace package body pachet_mascare is
    type tip_tabind is table of number index by pls_integer;
    v_tabind tip_tabind;

    function f_mascare(sir varchar2) return varchar2 is
        v_sir varchar2(100);
        v_lung number;
    begin
        v_sir := substr(sir,1,1);
        select length(sir) into v_lung from dual;
        v_sir := rpad(v_sir,v_lung,'*'); -- se pastreaza doar
        -- prima litera si apoi se pun stelute pana

```

```

        -- la lungimea sirului original
        return v_sir;
    end f_mascare;

function f_mascare(nr number) return number is
    lung number;
    minnou number;
    maxnou number;
    l_seed VARCHAR2(100);
    v_nrnou number;
begin
    if v_tabind.exists(nr) then
        return v_tabind(nr); --va fi folosit cand mascam cheia
                                --externa din tabelul REZOLVA
    else
        --generam un numar random care sa inceapa cu aceeasi
        --cifra ca nr si sa aiba aceeasi lungime
        lung:=length(to_char(nr));
        dbms_output.put_line('lung='||lung);
        minnou:=to_number(rpad(substr(to_char(nr),1,1),lung,'0'));
        maxnou:=to_number(rpad(substr(to_char(nr),1,1),lung,'9'));
        dbms_output.put_line('minnou='||minnou||' maxnou=' ||
                                maxnou);

        l_seed:=TO_CHAR(SYSTIMESTAMP,'YYYYDDMMHH24MISSFFFF');
        DBMS_RANDOM.seed (val => l_seed);

        v_nrnou:=round(DBMS_RANDOM.value(
                                low => minnou, high => maxnou),0);
        v_tabind(nr):=v_nrnou;
        return v_nrnou;
    end if;
end f_mascare;

function f_mascaregrupa(nr number) return number is
    v_nrnou number;
    lung number;
begin
    lung:=length(to_char(nr));
    v_nrnou:=to_number(rpad(substr(to_char(nr),1,1),lung,'0'));
    return v_nrnou;
end;

end;
/

```

- **Pasul 2)** Export al tabelelor dorite aplicând funcția de mascare din pachet

- În prealabil vom crea un director :

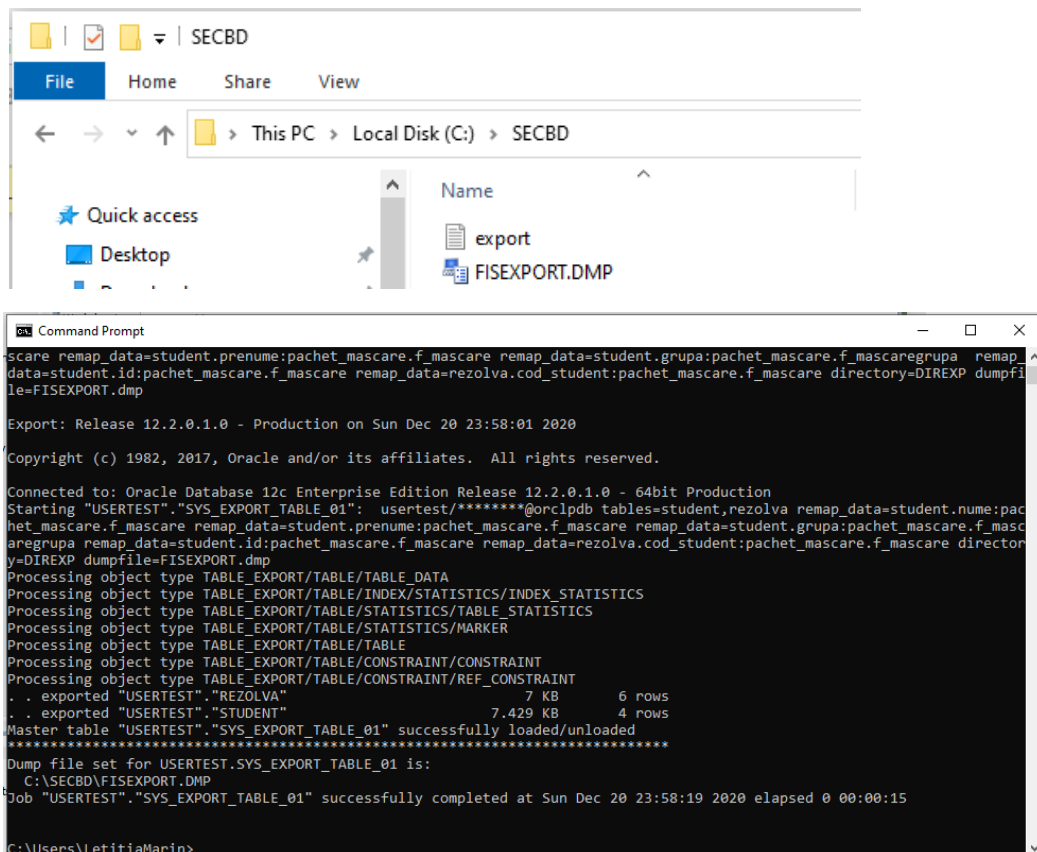
```
CREATE OR REPLACE DIRECTORY DIREXP AS 'C:\SECBDB';
```

- Din linie de comandă executăm comanda de export de date cu mascare:

```
expdp usertest/usertest@orclpdb tables=student,rezolva
remap_data=student.numa:pachet_mascare.f_mascare
remap_data=student.prenume:pachet_mascare.f_mascare
remap_data=student.grupa:pachet_mascare.f_mascaregrupaa
remap_data=student.id:pachet_mascare.f_mascare
remap_data=rezolva.cod_student:pachet_mascare.f_mascare
directory=DIREXP dumpfile=FISEXPORT.dmp
```

Remarcăm faptul că NU se va pune punct și virgulă la sfârșitul comenzii de export.

- Rezultatul se va găsi în directorul C:/SECBDB:



- **Pasul 3) Import al tabelelor cu date mascate (comanda NU are „;” !)**

```
impdp usertest/usertest@orclpdb directory=DIREXP
dumpfile=FISEXPORT.DMP
TABLES=student,rezolva remap_table=student:stud1
remap_table=rezolva:rez1
```

```

C:\Users\LetitiaMarin>impdp userestest/userestest@orclpdb directory=DIREXP dumpfile=FIEXPORT.DMP TABLES=student,rezolva remap_table=student:stud1 remap_table=rezolva:rez1

Import: Release 12.2.0.1.0 - Production on Mon Dec 21 00:02:15 2020

Copyright (c) 1982, 2017, Oracle and/or its affiliates. All rights reserved.

Connected to: Oracle Database 12c Enterprise Edition Release 12.2.0.1.0 - 64bit Production
Master table "USERTEST"."SYS_IMPORT_TABLE_01" successfully loaded/unloaded
Starting "USERTEST"."SYS_IMPORT_TABLE_01": userestest/*****@orclpdb directory=DIREXP dumpfile=FIEXPORT.DMP TABLES=student,rezolva remap_table=student:stud1 remap_table=rezolva:rez1
Processing object type TABLE_EXPORT/TABLE/TABLE
Processing object type TABLE_EXPORT/TABLE/TABLE_DATA
.. imported "USERTEST"."REZ1" 7 KB 6 rows
.. imported "USERTEST"."STUD1" 7.429 KB 4 rows
Processing object type TABLE_EXPORT/TABLE/CONSTRAINT/CONSTRAINT
Processing object type TABLE_EXPORT/TABLE/INDEX/STATISTICS/INDEX_STATISTICS
Processing object type TABLE_EXPORT/TABLE/CONSTRAINT/REF_CONSTRAINT
Processing object type TABLE_EXPORT/TABLE/STATISTICS/TABLE_STATISTICS
Processing object type TABLE_EXPORT/TABLE/STATISTICS/MARKER
Job "USERTEST"."SYS_IMPORT_TABLE_01" successfully completed at Mon Dec 21 00:02:25 2020 elapsed 0 00:00:09

C:\Users\LetitiaMarin>

```

Comanda SELECT pe tabelele STUD1 și REZ1 va afișa datele mascate importate:

SQL> select * from stud1;

ID	NUME	PRENUME	ANUL	SPE	GRUPA
164	A*****	A****	5	Inf	500
264	A****	T****	5	Inf	500
364	T****	A**	5	Inf	500
460	C*****	A*****	5	Inf	500

SQL> select * from rez1;

COD_STUDENT	COD_TEMA	DATA_UPLO	NOTA	DATA_CORE
164	1	04-NOV-12		
264	1	09-NOV-12		
164	2	19-NOV-12		
264	2	19-NOV-12		
364	2	24-NOV-12		
164	3	24-DEC-12		

6 rows selected.

Bibliografie:

<https://docs.oracle.com/en/database/oracle/oracle-database/12.2/sutil/oracle-data-pump.html#GUID-501A9908-BCC5-434C-8853-9A6096766B5A>

<http://www.oracle-base.com/articles/10g/oracle-data-pump-10g.php#GettingStarted>