

Laborator 6. Mascarea datelor in Oracle

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

Când sunt instalate în sistemul de producție, aplicațiile se execută pe date reale. Însa, 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.

Însa 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, ș.a.

Se pune problema mascării datelor confidențiale.

Există mai multe instrumente furnizate de Oracle (Oracle Data Masking Pack, Oracle Data Pump, etc) și există de asemenea posibilitatea implementării unor subprograme personalizate 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 că 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, să se asigure modificarea corespunzătoare în cascada 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 tabelă.

În continuare se va prezenta metode 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 Oracle (ex: C:\app2\mm\product\11.2.0\dbhome_1\BIN). Alternativ, se pot crea job-uri folosind API specific¹.

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

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

expdp user/parola tables=tabela1,tabela2,...

remap_data=tabela1.colx:pachet.funcție_de_mascare

remap_data=tabela2.coly:pachet.funcție_de_mascare2 ... directory=NUMEDIR

dumpfile=NUMEFIS.dmp

Pasul 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 ...

¹ DATA PUMP API , <http://www.oracle-base.com/articles/10g/oracle-data-pump-10g.php#GettingStarted>

Exercițiu:

Se considera tabelele STUDENT si REZOLVA cu următoarea structura:

```
SQL> desc student;
Name                Null?    Type
-----
ID                   NOT NULL NUMBER
NUME                 UARCHAR2(30)
PRENUME              UARCHAR2(30)
ANUL                 NUMBER
SPECIALIZARE         UARCHAR2(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
```

Sa se utilizeze mascarea datelor pentru coloanele *id*, *nume*, *prenume*, *grupa* din tabela student. Se cunoaște următorul fapt: *cod_student* din tabela REZOLVA este cheie externa si refera cheia primara (*id*) din tabela STUDENT.

Mențiune:

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	Tudor	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 in care sa includem funcții de mascare a datelor

Se dorește păstrarea primului caracter din *nume* si din *prenume*, iar pentru *id* si *grupa* păstrarea primei cifre. Vom folosi o funcție supraîncărcată in pachet pentru siruri , respectiv pentru cheia primara *id* , si o alta funcție pentru mascare grupa.

Întrucât *id* este cheia primara referita, vom folosi un tablou indexat care sa tina pe poziția *id* valoarea mascata. Același element din tablou va fi folosit si 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,'*'); -- sa pastreze doar prima litera si
        apoi sa puna 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 REZOLVA
```

```
        else
```

```
            --generam un nr random care sa inceapa cu aceeasi cifra ca si 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 tabelor dorite aplicând funcția de mascare din pachet

In prealabil vom crea un director :

```
CREATE OR REPLACE DIRECTORY DIREXP7 AS 'C:\SECB D';
```

Din linie de comanda executam comanda de export de date cu mascare:

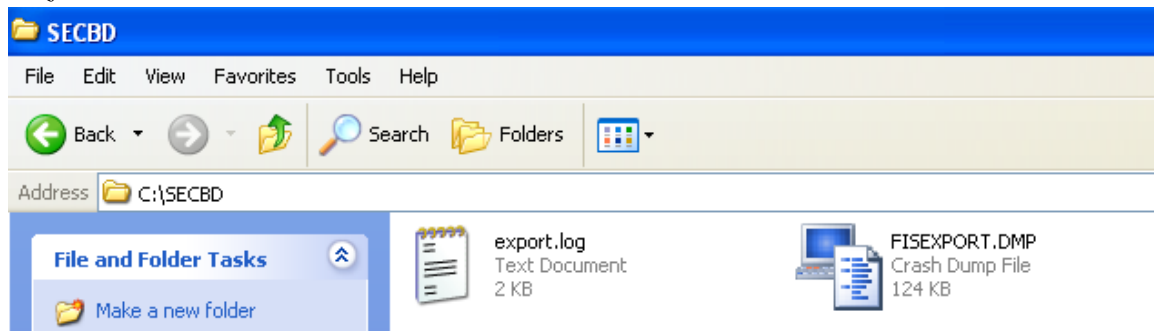
```

expdp camelia/12345 tables=student,rezolva
remap_data=student.num e:pachet_mascare.f_mascare
remap_data=student.prenume:pachet_mascare.f_mascare
remap_data=student.grupa:pachet_mascare.f_mascaregrupa
remap_data=student.id:pachet_mascare.f_mascare
remap_data=rezolva.cod_student:pachet_mascare.f_mascare
directory=DIREXP7 dumpfile=FISEXPORT.dmp

```

Remarcam faptul ca NU se va pune punct si virgula la sfarsitul comenzii de export.

Rezultatul in calea directorului C:/SECB D



```

C:\app2\mm\product\11.2.0\dbhome_1\BIN>expdp camelia/12345          tables=student,
rezolva remap_data=student.num:pachet_mascare.f_mascare remap_data=student.pr
enume:pachet_mascare.f_mascare remap_data=student.grupa:pachet_mascare.f_masc
aregrupa remap_data=student.id:pachet_mascare.f_mascare remap_data=rezolva.
cod_student:pachet_mascare.f_mascare directory=DIREXP7 dumpfile=FISEXPORT.dmp

Export: Release 11.2.0.1.0 - Production on Mon Dec 24 05:20:49 2012

Copyright (c) 1982, 2009, Oracle and/or its affiliates. All rights reserved.

Connected to: Oracle Database 11g Enterprise Edition Release 11.2.0.1.0 - Produc
tion
With the Partitioning, OLAP, Data Mining and Real Application Testing options
Starting "CAMELIA"."SYS_EXPORT_TABLE_01": camelia/***** tables=student,rezol
va remap_data=student.num:pachet_mascare.f_mascare remap_data=student.prenume:p
achet_mascare.f_mascare remap_data=student.grupa:pachet_mascare.f_mascaregrupa r
emap_data=student.id:pachet_mascare.f_mascare remap_data=rezolva.cod_student:pac
het_mascare.f_mascare directory=DIREXP7 dumpfile=FISEXPORT.dmp;
Estimate in progress using BLOCKS method...
Processing object type TABLE_EXPORT/TABLE/TABLE_DATA
Total estimation using BLOCKS method: 128 KB
Processing object type TABLE_EXPORT/TABLE/TABLE
Processing object type TABLE_EXPORT/TABLE/INDEX/INDEX
Processing object type TABLE_EXPORT/TABLE/CONSTRAINT/CONSTRAINT
Processing object type TABLE_EXPORT/TABLE/CONSTRAINT/REF_CONSTRAINT
. . exported "CAMELIA"."REZOLVA" 6.773 KB 6 rows
. . exported "CAMELIA"."STUDENT" 7.125 KB 4 rows
Master table "CAMELIA"."SYS_EXPORT_TABLE_01" successfully loaded/unloaded
*****
Dump file set for CAMELIA.SYS_EXPORT_TABLE_01 is:
C:\SECB\FISEXPORT.DMP;
Job "CAMELIA"."SYS_EXPORT_TABLE_01" successfully completed at 05:20:56

C:\app2\mm\product\11.2.0\dbhome_1\BIN>_

```

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

**impdp camelia/12345 directory=DIREXP7 dumpfile=FISEXPORT.DMP
TABLES=student,rezolva remap_table=student:stud1 remap_table=rezolva:rez1**

```

C:\app2\mm\product\11.2.0\dbhome_1\BIN>impdp camelia/12345          directory=DIREX
P7 dumpfile=FISEXPORT.DMP TABLES=student,rezolva remap_table=student:stud1 remap
_table=rezolva:rez1

Import: Release 11.2.0.1.0 - Production on Mon Dec 24 05:58:01 2012

Copyright (c) 1982, 2009, Oracle and/or its affiliates. All rights reserved.

Connected to: Oracle Database 11g Enterprise Edition Release 11.2.0.1.0 - Produc
tion
With the Partitioning, OLAP, Data Mining and Real Application Testing options
Master table "CAMELIA"."SYS_IMPORT_TABLE_01" successfully loaded/unloaded
Starting "CAMELIA"."SYS_IMPORT_TABLE_01": camelia/***** directory=DIREXP7 du
mpfile=FISEXPORT.DMP TABLES=student,rezolva remap_table=student:stud1 remap_tabl
e=rezolva:rez1
Processing object type TABLE_EXPORT/TABLE/TABLE
Processing object type TABLE_EXPORT/TABLE/TABLE_DATA
. . imported "CAMELIA"."REZ1" 6.773 KB 6 rows
. . imported "CAMELIA"."STUD1" 7.125 KB 4 rows
Processing object type TABLE_EXPORT/TABLE/INDEX/INDEX
ORA-31684: Object type INDEX:"CAMELIA"."PK_REZ1" already exists
Processing object type TABLE_EXPORT/TABLE/CONSTRAINT/CONSTRAINT
ORA-31684: Object type CONSTRAINT:"CAMELIA"."PK_REZ1" already exists
ORA-31684: Object type CONSTRAINT:"CAMELIA"."SYS_C0014783" already exists
Processing object type TABLE_EXPORT/TABLE/CONSTRAINT/REF_CONSTRAINT
ORA-31684: Object type REF_CONSTRAINT:"CAMELIA"."FK_REZ2" already exists
ORA-31684: Object type REF_CONSTRAINT:"CAMELIA"."FK_REZ1" already exists
Job "CAMELIA"."SYS_IMPORT_TABLE_01" completed with 5 error(s) at 05:58:05

C:\app2\mm\product\11.2.0\dbhome_1\BIN>

```

Select din tabelele STUD1 si REZ1 arata 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:

http://download.oracle.com/otndocs/products/database/enterprise_edition/utilities/pdf/datapump11g2_transform_1009.pdf

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