

Wie sehen Ihr Testdaten aus ? - Anonymisierung mit Oracle Data Masking

Dirk Braunecker
Logica Deutschland GmbH & Co. KG, now part of CGI
Sulzbach (Taunus)

Schlüsselworte

Datenbank, Testdaten, Anonymisierung

Einleitung

Im Zuge des Projektfortschritts entsteht früher oder später das Problem, Testdaten zu organisieren und diese ggf. auch zu anonymisieren.

Des Weiteren dürfen in vielen Unternehmen aufgrund datenschutzrechtlicher Einschränkungen Produktivdaten nur anonymisiert als Testdaten verwendet werden. Die Handhabung und Administration der Anonymisierung sollte daher einfach und flexibel sein, um auch evtl. Anpassungen schnell umzusetzen. Oracle bietet mit Data Masking ein Werkzeug zur Anonymisierung von Daten zur weiteren Verwendung z. B. in Entwicklungs- und Testumgebungen.

Der Vortrag stellt das Werkzeug Oracle Data Masking vor und zeigt die Einsatzmöglichkeiten des Tools auf.

Testdaten

In nahezu allen Projekten tritt früher oder später die Frage nach den Testdaten auf. Diese Testdaten können aus folgenden Quellen stammen:

- Testsystem
- Entwicklersystem
- Produktionssystem
- Archiv

Je nach Ursprung der Testdaten ist eine gewisse Sensibilität mit dem Umgang der Daten notwendig. Oftmals werden Daten 1:1 direkt aus dem Produktionssystem verwendet, ohne vorherige Prüfung auf Relevanz der Anonymisierung. Vor allem in hochsensiblen Umgebungen wie z. B. Banken und Versicherung ist es absolut zwingend diese Daten zu anonymisieren, sofern diese dem Produktionssystem entsprungen. Daraus aufbauend sind folgende Entscheidungen zu treffen:

- Wo kommen die Testdaten her ?
- Welche sensible Daten sind enthalten ?

Diese Fragen/Anforderungen und deren Antworten/Umsetzungen sind in einem Testdatenkonzept zu integrieren und müssen den Entwicklungsteams zur Verfügung gestellt werden.

Die weitere Aufgabe besteht nun darin festzulegen, wie sensible Daten anonymisiert werden können.

Die sensiblen Daten werden z. B. mit folgenden Methoden anonymisiert:

- Ersetzung durch Konstanten
- Verfremdung durch einen Algorithmus
- Löschen von Teilen der Daten
- Ersetzen durch Zufallswerte

Ein oft angewendeter Ansatz ist das manuelle Klonen der Daten in separate Tabellen mit anschließender Manipulation der sensiblen Daten mittels SQL-Befehlen oder PL/SQL-Quellcode. Der hierbei entstehende administrative Aufwand ist recht hoch und wird in der Regel auch manuell durchgeführt.

Ein besserer weil einfacher Ansatz ist die toolbasierte Anonymisierung von Daten. Oracle bietet hierfür das Tool Oracle Data Masking an. Dieses Tool ist im Enterprise Manager integriert und muss separat lizenziert werden (9077 € pro Prozessor + Supportkosten).

Oracle Data Masking

Das Tool ist wie bereits erwähnt im Oracle Enterprise Manager integriert und enthält folgende Features:

- Welche Spalten sollen anonymisiert werden ?
- Wie sollen die Spalten anonymisiert werden ?
- Sind Fremdschlüssel davon betroffen ?
- Sollen die anonymisierten Daten in eine separate Datenbank geklont werden ?

Die Administration des Data Masking Tools befindet sich im Reiter „Schema“ des Enterprise Managers

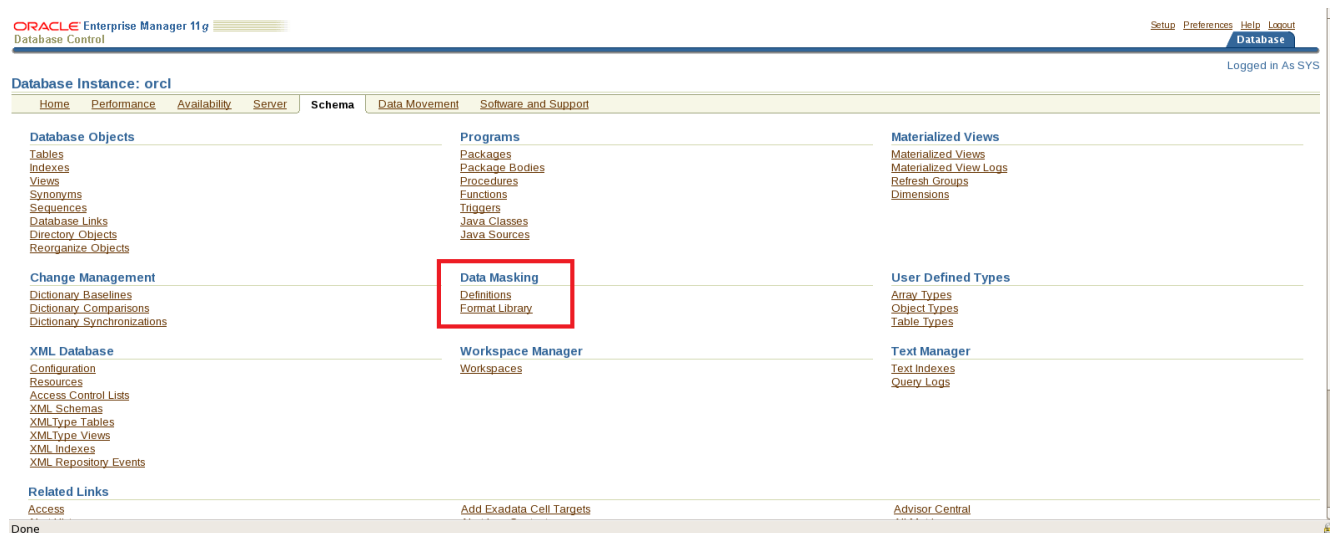


Abb. 1: Administration

In der Menüoption „Definitions“ können Data Masking Definitionen durchgeführt werden.

Confirmation
Masking Definition Salary has been deleted successfully.

Data Masking Definitions
Data masking is the process of making sensitive information in test or non-production databases safe. It disguises sensitive information by overwriting it with realistic looking but false data of a similar type. A masking definition defines the columns to be masked and the format of masked data. You can create a new masking definition or use an existing definition for a masking operation. The Format Library contains a collection of ready-to-use masking formats.

Search Database "orcl"

Select Masking Definition	Database	Description	Columns	Status	Most Recent Job Ended
No definitions					

Format Library
A masking format defines the format of masked data. You can create a new masking format and reuse it later when creating a masking definition.
[Format Library](#)

[Database](#) | [Setup](#) | [Preferences](#) | [Help](#) | [Logout](#)

Copyright © 1996, 2010, Oracle. All rights reserved.

Abb. 2: Definitionen

Nach Ausführen des Buttons „Create“ wird eine neue Definition erzeugt

ORACLE Enterprise Manager 11g Database Control [Setup](#) [Preferences](#) [Help](#) [Logout](#) **Database**

Create Masking Definition

* Name SALARY_DEFINITION
* Database orcl
Description Anonymisierung des Gehalts

Columns
Add columns you want to mask and define masking format for each column. Foreign key columns are automatically added to maintain referential integrity. Dependent columns are columns that do not have foreign key constraints defined, but reference a masked column due to application level constraints. You can manually add dependent columns to a masked column. Removing a column from this list will remove all foreign key and dependent columns.

Select	Owner	Table	Column	Column Group	Data Type	Format	Foreign Key Columns	Dependent Columns
							Count	Add
No columns added								

Foreign Key Columns

Owner	Table	Column	Parent Owner	Parent Table	Parent Column
No foreign key columns					

Abb.3: Definitionen anlegen

Im Feld Name und Description wurden der Name der Definition sowie deren Beschreibung eingegeben. Mit dem Button „Add“ werden anschließend die Details der Definition festgelegt. Im ersten Schritt werden die Tabelle und die Spalten ausgewählt.

ORACLE Enterprise Manager 11g Database Control Setup Preferences Help Logout

Data Masking Definitions > Create Masking Definition >

Add Columns

Database **orcl** Logged In As **SYS** Cancel Add Define Format And Add

Add one or more columns for masking. Foreign key columns will be added automatically. You can define masking format at once for all selected columns if they have the same data type.

Search

Schema Column Name
 Table Name Column Comment
Enter a string in column comments.

Mask selected columns as a group

Select	Owner	Table Name	Column Name	Data Type	Comment
No	columns				

Abb.4: Tabelle und Spalten auswählen

Anschließend werden über den Button „Define Format and Add“ die Formatierung/Anonymisierung der Daten festgelegt.

Data Masking Definitions > Create Masking Definition >

Information
 The following columns have check constraints other than NOT NULL. Make sure that the defined masking formats satisfy check conditions.
[HR.EMPLOYEES.SALARY](#) - salary > 0

Define Column Mask

Owner **HR** Table **EMPLOYEES** Cancel OK
 Column **SALARY** Data Type **NUMBER(8,2)**

By default all records in the table will be masked using the specified format. You can optionally identify more than one subset of records using conditions. Each subset can be masked using a corresponding masking format. The subsets will be masked in the order they are specified. A subset will not be masked again even when it matches a subsequent condition.

Format Entry

|

Select	Condition	Format Entry Properties				Sample	Remove
		Property	Value	Property	Value		
<input type="checkbox"/>	Conditions						
<input checked="" type="radio"/>	Default Condition (Add a format entry)						

Abb.5: Format auswählen

Im Format Entry-Feld wurde Fixed Number = Konstante ausgewählt und mit dem Button „Add“ hinzugefügt, die Konstante eingeben und anschließend mit „Ok“ bestätigt.

Data Masking Definitions > Create Masking Definition >

Define Column Mask

Owner **HR** Table **EMPLOYEES** Cancel OK
 Column **SALARY** Data Type **NUMBER(8,2)**

By default all records in the table will be masked using the specified format. You can optionally identify more than one subset of records using conditions. Each subset can be masked using a corresponding masking format. The subsets will be masked in the order they are specified. A subset will not be masked again even when it matches a subsequent condition.

Add Condition

Import Format Format Entry **Fixed Number** Add

Expand All | Collapse All

Select	Condition	Format Entry Properties				Sample	Remove
		Property	Value	Property	Value		
<input type="radio"/>	Conditions						
<input type="radio"/>	Default Condition						
<input type="radio"/>	Fixed Number	Fixed Number	<input type="text" value="4711"/>				

Cancel OK

[Database](#) | [Setup](#) | [Preferences](#) | [Help](#) | [Logout](#)

Abb.6: Format hinzugefügt

Abschließend wird die Definition folgendermaßen angezeigt:

ORACLE Enterprise Manager 11g Database Control [Setup](#) [Preferences](#) [Help](#) [Logout](#)

Data Masking Definitions

Data masking is the process of making sensitive information in test or non-production databases safe. It disguises sensitive information by overwriting it with realistic looking but false data of a similar type. A masking definition defines the columns to be masked and the format of masked data. You can create a new masking definition or use an existing definition for a masking operation. The Format Library contains a collection of ready-to-use masking formats.

Search Go Import Create

View Edit Create Like Generate Script Schedule Job Export Clone Database Delete

Select	Masking Definition	Database	Description	Columns	Status	Most Recent Job Ended
<input checked="" type="radio"/>	SALARY_DEFINITION	orcl	Anonymisierung des Gehalts		1 Script Not Generated	

Format Library
 A masking format defines the format of masked data. You can create a new masking format and reuse it later when creating a masking definition.
[Format Library](#)

Abb. 6: Definition erstellt

Danach sind noch zwei Arbeitsschritte notwendig, um die Anonymisierung zu aktivieren

- SQL-Skript erzeugen (Button „Generate Script“)
- Job-Ausführung zur Aktivierung (Button „Schedule Job“)

Danach kann mittels des SQL-Worksheets der Spalten-Inhalt selektiert werden, der nun anonymisiert ausgegeben werden muss. In der Abbildung ist klar ersichtlich, dass der ursprüngliche Wert im Feld SALARY nun mit der definierten Konstante 4711 ersetzt wird.

Last Executed SQL

```
SELECT * FROM hr.employees
```

Last Execution Details

[SQL Repair Advisor](#) [SQL Details](#) [Schedule SQL Tuning Advisor](#)

Results										
Statistics										
Plan										
Execution Time (seconds) 0.029										
EMPLOYEE_ID	FIRST_NAME	LAST_NAME	EMAIL	PHONE_NUMBER	HIRE_DATE	JOB_ID	SALARY	COMMISSION_PCT	MANAGER_ID	DEPARTMENT_ID
198	Donald	OConnell	DOCONNEL	650.507.9833	Jun 21, 1999 12:00:00 AM	SH_CLERK	4711		124	50
199	Douglas	Grant	DGRANT	650.507.9844	Jan 13, 2000 12:00:00 AM	SH_CLERK	4711		124	50
200	Jennifer	Whalen	JWHALEN	515.123.4444	Sep 17, 1987 12:00:00 AM	AD_ASST	4711		101	10
201	Michael	Hartstein	MHARTSTE	515.123.5555	Feb 17, 1996 12:00:00 AM	MK_MAN	4711		100	20
202	Pat	Fay	PFAY	603.123.6666	Aug 17, 1997 12:00:00 AM	MK_REP	4711		201	20
203	Susan	Mavris	SMAVRIS	515.123.7777	Jun 7, 1994 12:00:00 AM	HR_REP	4711		101	40
204	Hermann	Baer	HBAER	515.123.8888	Jun 7, 1994 12:00:00 AM	PR_REP	4711		101	70
205	Shelley	Higgins	SHIGGINS	515.123.8080	Jun 7, 1994 12:00:00 AM	AC_MGR	4711		101	110
206	William	Gietz	WGIEZT	515.123.8181	Jun 7, 1994 12:00:00 AM	AC_ACCOUNT	4711		205	110
100	Steven	King	SKING	515.123.4567	Jun 17, 1987 12:00:00 AM	AD PRES	4711			90
101	Neena	Kochhar	NKOCHHAR	515.123.4568	Sep 21, 1989 12:00:00 AM	AD_VP	4711		100	90
102	Lex	De Haan	LDEHAAN	515.123.4569	Jan 13, 1993 12:00:00 AM	AD_VP	4711		100	90
103	Alexander	Hunold	AHUNOLD	590.423.4567	Jan 3, 1990 12:00:00 AM	IT_PROG	4711		102	60
104	Bruce	Ernst	BERNST	590.423.4568	May 21, 1991 12:00:00 AM	IT_PROG	4711		103	60
105	David	Austin	DAUSTIN	590.423.4569	Jun 25, 1997 12:00:00 AM	IT_PROG	4711		103	60
106	Valli	Pataballa	VPATABAL	590.423.4560	Feb 5, 1998 12:00:00 AM	IT_PROG	4711		103	60
107	Diana	Lorentz	DLORENTZ	590.423.5567	Feb 7, 1999 12:00:00 AM	IT_PROG	4711		103	60

Abb.7:Anonymisierung

Die Anonymisierung mit Oracle Data Masking kann in folgenden Schritten durchgeführt werden:

1. Testdaten-Datenbank erstellen (evtl. durch Klonen)
2. Data Masking Definitionen erstellen und ausführen/aktivieren
3. Datenbank als Entwicklungs- oder Testdatenbank klonen

Dadurch entsteht eine Master-Datenbank die als Grundlage für alle Entwicklungs- und Testdatenbanken dienen kann. Die Definitionen werden zentral nur in einer Datenbank gepflegt und administriert und anschließend den Projekten zur Verfügung gestellt.

Fazit

Grundsätzlich wird empfohlen eine Strategie im Unternehmen oder Projekt zu entwickeln, die genau festlegt, wie und wann Testdaten anonymisiert werden müssen. Diese Vorgaben sollten im Rahmen eines Prozesses oder Leitfadens genau festlegen, welche Daten bzw. Datenstrukturen für die Anonymisierung vorzusehen sind. Oracle Data Masking ist hierbei eine einfache und effektive Alternative zur Durchführung der Anonymisierung von Testdaten.

Kontaktadresse:

Dirk Braunecker

Logica Deutschland GmbH & Co. KG | Now part of CGI

Am Limespark 2

D-65843 Sulzbach (Taunus)

Telefon: +49 151 4223 1010

E-Mail dirk.braunecker@cgi.com | dirk.braunecker@cgi.com

Internet: www.logica.de | www.cgi.com