Protecting Privacy with Data Masking
Oracle Enterprise Manager 10gR4

Nicholas J. Donatone
Principal Grid Sales Consultant
Grid Control - Infrastructure Management
Broad management for all application-dependent components

Extended Infrastructure Management

Virtual Servers
- VMWare
- Oracle VM

Operating Systems
- Linux
- All Unix
- Microsoft

Servers
- Dell PowerEdge

Databases
- Oracle
- IBM DB2
- MS SQL Server (2000 & 2005)
- Sybase

Middleware
- Oracle
- IBM WebSphere
- BEA WebLogic
- MS .NET
- MS BizTalk Server
- MS IIS
- MS Commerce Server
- MS ISA Server
- MS Active Directory
- J Boss AS
- IBM MQ Series
- Citrix Presentation Server (partner-built)
- Blue Lane PatchPoint (partner-built)
- Tomcat

Network
- Check Point Firewall
- Juniper Netscreen Firewall
- F5 BIG-IP
- Nortel Alteon Switch (partner-built)

Storage
- NetApp Filer
- EMC Celerra, Clariion, Symmetrix
- Pillar Axiom (partner-built)
- Onaro SANScreen (partner-built)

Applications
- Oracle
- Microsoft Exchange
- SAP
Business Drivers for Data sharing

**Application Testing**
- Offshore or in-house application development
- Offshore or in-house software QA

**Data sharing**
- Claims processing
- Offline reporting
- Data archival and retrieval
- Marketing analysis of customer data

- California Database Security Breach Notification Act
- Sarbanes Oxley Act of 2002
- US HIPAA Act
- US Graham-Leach-Bliley Financial Services Modernization Act
- EU Data Protection Directive
Data Masking
Data masking concepts

What
- The act of anonymizing customer, financial, or company confidential data to create new, legible data which retains the data's properties, such as its width, type, and format.

Why
- To protect confidential data in test environments when the data is used by developers or offshore vendors
- When customer data is shared with 3rd parties without revealing personally identifiable information

<table>
<thead>
<tr>
<th>LAST_NAME</th>
<th>SSN</th>
<th>SALARY</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGUI LAR</td>
<td>203-33-3234</td>
<td>40,000</td>
</tr>
<tr>
<td>BENSON</td>
<td>323-22-2943</td>
<td>60,000</td>
</tr>
<tr>
<td>D' SOUZA</td>
<td>989-22-2403</td>
<td>80,000</td>
</tr>
<tr>
<td>FI ORANO</td>
<td>093-44-3823</td>
<td>45,000</td>
</tr>
<tr>
<td>ANSKEKSL</td>
<td>111-23-1111</td>
<td>40,000</td>
</tr>
<tr>
<td>BKJHHEIEDK</td>
<td>111-34-1345</td>
<td>60,000</td>
</tr>
<tr>
<td>KDDEHLHESA</td>
<td>111-97-2749</td>
<td>80,000</td>
</tr>
<tr>
<td>FPENZXIEK</td>
<td>111-49-3849</td>
<td>45,000</td>
</tr>
</tbody>
</table>
Data Masking

Major features

- Automatic database referential integrity when masking primary keys
  - Implicit – database enforced
  - Explicit – application enforced
- Data mask format library
- View sample data before masking
- Application masking templates
- Define once; execute multiple times
Format Libraries

- Mask Primitives
  - Random Number
  - Random String
  - Random Date within range
  - Shuffle
  - Sub string of original value
  - Table Column
- User Defined Function
  - National Identifiers
  - Social Security Numbers
  - Credit Card Numbers
User-defined mask formats

Email notification testing

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>User Defined Function</td>
<td>Function Name: hr.email_mask</td>
</tr>
</tbody>
</table>

**Sample Masked Data**

Samples are generated using defined format. Use Refresh to re-generate samples.

- Harry Dean.978448089.Dvree@spambob.com
- Daniel.118074087.Bates@spambob.com
- Dieter.421172061.Elliott@spambob.com
- Mani.268599080.Field@spambob.com
- Gena.417785060.Altman@spambob.com
Masking Definitions

- Associates formats with database
  - Maps formats to table columns being masked
  - Defines dependent columns
  - Associated Database target
- Automatically identifies Foreign key relationships
- Can specify undeclared constraints as related columns
- Import-from or export-to XML
- “Create like” to apply to similar databases
Referential Integrity Enforcement

**Masking Definition: Employee HR mask**

A data masking definition specifies what columns to be masked and the format of masked data.

<table>
<thead>
<tr>
<th>Name</th>
<th>Employee HR mask</th>
<th>Database</th>
<th>Oemrep_Database</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Columns**

<table>
<thead>
<tr>
<th>Owner</th>
<th>Table Name</th>
<th>Column Name</th>
<th>Format</th>
</tr>
</thead>
<tbody>
<tr>
<td>HR</td>
<td>EMPLOYEES</td>
<td>EMPLOYEE_ID</td>
<td></td>
</tr>
<tr>
<td>HR</td>
<td>EMPLOYEES</td>
<td>FIRST_NAME</td>
<td></td>
</tr>
<tr>
<td>HR</td>
<td>EMPLOYEES</td>
<td>LAST_NAME</td>
<td></td>
</tr>
</tbody>
</table>

**Foreign Key Columns**

<table>
<thead>
<tr>
<th>Owner</th>
<th>Table Name</th>
<th>Column Name</th>
<th>Parent Owner</th>
<th>Parent Table</th>
<th>Parent Column</th>
</tr>
</thead>
<tbody>
<tr>
<td>HR</td>
<td>DEPARTMENTS</td>
<td>MANAGER_ID</td>
<td>HR</td>
<td>EMPLOYEES</td>
<td>EMPLOYEE_ID</td>
</tr>
<tr>
<td>HR</td>
<td>EMPLOYEES</td>
<td>MANAGER_ID</td>
<td>HR</td>
<td>EMPLOYEES</td>
<td>EMPLOYEE_ID</td>
</tr>
<tr>
<td>HR</td>
<td>JOB_HISTORY</td>
<td>EMPLOYEE_ID</td>
<td>HR</td>
<td>EMPLOYEES</td>
<td>EMPLOYEE_ID</td>
</tr>
<tr>
<td>OE</td>
<td>CUSTOMERS</td>
<td>ACCOUNT_MGR_ID</td>
<td>HR</td>
<td>EMPLOYEES</td>
<td>EMPLOYEE_ID</td>
</tr>
<tr>
<td>OE</td>
<td>ORDERS</td>
<td>SALES_REP_ID</td>
<td>HR</td>
<td>EMPLOYEES</td>
<td>EMPLOYEE_ID</td>
</tr>
</tbody>
</table>

**Dependent Columns**

<table>
<thead>
<tr>
<th>Owner</th>
<th>Table Name</th>
<th>Column Name</th>
<th>Parent Owner</th>
<th>Parent Table</th>
<th>Parent Column</th>
</tr>
</thead>
<tbody>
<tr>
<td>HR</td>
<td>MANAGERS</td>
<td>MGR_ID</td>
<td>HR</td>
<td>EMPLOYEES</td>
<td>EMPLOYEE_ID</td>
</tr>
</tbody>
</table>

*Database-enforced*

*Application-enforced*
Pre-Masking Validation

- Ensure uniqueness can be maintained
- Ensure formats match column data types
- Check Space availability
- Warn about Check Constraints
- Check presence of default Partitions
Masking Workflow

Security Admin:
- Identify Sensitive Information
- Identify Data Formats
- Format Library
- Masking Definition

DBA:
- Clone Prod to Staging
- Review Mask Definition
- Execute Mask
- Clone Staging to Test

Prod -> Staging -> Test
Data Masking Internals

Build Mapping table
   Contains orig_value, mask_value)

Disable Constraints

Rename table

Recreate & Reload from renamed table and mapping tables

Drop Renamed table and mapping table

Collect statistics

Enable Constraints
Performance

- Optimizations
  - SQL Parallelism for tables > 1 million rows
  - Statistics collection before & after masking
  - CTAS statement with NOLOGGING

- Test results
  - Case 1
    - 60GB Database
    - 100 tables, 215 columns
    - 20mins
  - Case 2
    - 6 column, 100 million row table
    - Random Number
    - 1.3 hours
For More Information

http://search.oracle.com

Data Masking
Demonstration
Data Masking

- Helps sharing of production data in compliance with data privacy policies.
- Delivers uniform application of mask formats across all enterprise data.
- Increase DBA productivity by automating the discovery and masking of sensitive data.