DBA Best Practices for Protecting Data Privacy with Oracle Data Masking

Waleed Ahmed, Principal Product Manager
Oracle
Tom Jernejcic, Supervising DBA
Riverside County
Safe Harbor Statement

The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, and timing of any features or functionality described for Oracle’s products remains at the sole discretion of Oracle.
Program Agenda

- Data Privacy Challenges
- Privacy in Database Cloud
  - Challenges and Solution
- Data Masking Best Practice for DBAs
  - Recommendations, Examples, At-Source Masking in Snapshot Standby
- New Features
- Riverside County, CA – Case Study
Data Privacy Challenges
Data Privacy Challenges

What high-risk data is collected?
Where is it stored?
How is it protected?

<table>
<thead>
<tr>
<th>Challenges</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Production</strong></td>
</tr>
<tr>
<td>• Database growth in the cloud</td>
</tr>
<tr>
<td>• Where does sensitive data reside</td>
</tr>
<tr>
<td>• Who has access to what?</td>
</tr>
<tr>
<td><strong>Development, Test, Training…</strong></td>
</tr>
<tr>
<td>• Full size database copy of production for non-production use</td>
</tr>
<tr>
<td>• Sensitive data exposed to unprivileged users</td>
</tr>
<tr>
<td>• Without realistic data, development and test is done out of context of production</td>
</tr>
<tr>
<td>• Traditional approach of protecting data is inaccurate, cumbersome and renders application unusable</td>
</tr>
</tbody>
</table>
Data Privacy Solution

Data Masking
• Replace sensitive production data with fictitious yet realistic data in Development and Test
• Comply with regulations while enabling realistic application testing

Data Subsetting
• Provision a fractional yet relationally intact representation from production data in Development
• Reduce storage costs while enabling realistic application development

Secure Application Testing
Securing Non-Production Databases
Data Masking and Subsetting Methodology

- **Find**: Locate and Identify sensitive data and discover referential relationships
- **Assess**: Define the optimal data masking formats and subset criteria
- **Secure**: Provision secured test / dev systems using masking + subsetting
- **Test**: Verify the integrity of applications and security of data

<table>
<thead>
<tr>
<th>DEPT</th>
<th>LAST_NAME</th>
<th>SSN</th>
<th>SALARY</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>AGUILAR</td>
<td>203-33-3234</td>
<td>40,000</td>
</tr>
<tr>
<td>20</td>
<td>BENSON</td>
<td>543-44-2737</td>
<td>30,000</td>
</tr>
<tr>
<td>10</td>
<td>CHANDRA</td>
<td>323-22-2943</td>
<td>60,000</td>
</tr>
<tr>
<td>30</td>
<td>DONNER</td>
<td>270-24-4023</td>
<td>50,000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DEPT</th>
<th>LAST_NAME</th>
<th>SSN</th>
<th>SALARY</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>SMITH</td>
<td>111—23—1111</td>
<td>40,000</td>
</tr>
<tr>
<td>10</td>
<td>JOHNSON</td>
<td>222-34-1345</td>
<td>60,000</td>
</tr>
</tbody>
</table>

IBM DB2
MS SQL
Privacy in Database Cloud
Today’s Forecast: Cloudy with a chance of breach

Cloud Data Sprawl

- More than 50% of Global companies will have customer-sensitive data stored in public cloud by 2016*
- Worldwide public cloud services sees explosive growth**
- Average cost of data breach was $5.5M per company in 2011***
Dev, Test, QA, Training database are experiencing cloud sprawl

Non-production databases are being provisioned rapidly and at full size

How do you locate sensitive databases and data in this rapidly growing arena and secure it?
Securing Non-Production Databases
Methodology – Find, Assess, Secure and Test

Find
- Discover and locate sensitive database and data

Assess
- Classify sensitive data
- Assign masking definitions and or subset definition to databases

Secure
- Execute mask and or subset and validate

Test
Securing Non-Production Databases

- Configure
  - Create sensitive column types
  - Create application data model
- Discovery
  - Run sensitive column discovery
- Classify
  - Create data masking definition
  - Create subset definition
- Execution
  - At-Source
  - In-Place
  - Heterogeneous
Securing Non-Production Databases

- Oracle Enterprise Manager allows you to identify databases that may contain sensitive data automatically in your environment in Oracle Enterprise Manager.
- Assigning an Application Data Model to the sensitive data discovered allows for masking and or subset definitions to be created.
Find: Securing Non-Production Databases – Classify

DBA evaluates results from sensitive column discovery
Assigns or creates an Application Data Model
Create data masking definition
Define masking format type
Define pre/post masking steps
Assess – Secure – Test Data Management

- Create subsetting definition
- Define
  - Table rules
  - Column rules
  - Rule parameters
  - Data masking
  - Pre/post subset script
Execute – Data Masking

Secure:
Execute the masking job using one of three ways:
1. In-Place Masking
   - Mask the data in place
2. At-Source Masking
   - Create masked export dump files that can be imported into non-production databases
3. Database Gateway
   - Mask data in non-oracle database

Test:
Validate the data is masked and is in a state that allows the application to function correctly
Securing Non-Production

Closing the Security Gap

- Data stores are expanding at a rate greater than 20% per year*
- Greatest risk, threats or vulnerability of data by human or internal users is over 70%*
- Oracle Enterprise Manager simplifies and automates the identification and protection of sensitive data in an enterprise
- Employing the F.A.S.T method with Oracle Enterprise Manager simplifies the discovery of sensitive data within an enterprise with ease

*IOUG 2012 – Closing the Security Gap
Data Masking Best Practice for DBA’s
Data Masking Best Practice

Modes of Data Masking

- **In-Place Data Masking**
  - Masking runs in the target database (staging) replacing production data with fictitious yet realistic data

- **At-Source Masking**
  - Production data is masked as it is being exported to binary files without affecting production data within the database

- **Heterogeneous Masking**
  - Data is masked in non-oracle database using Oracle Database Gateway

Each of the above methods requires correct environment configurations to achieve the best performance
In-Place Data Masking Best Practice

Correct hardware configuration
- I/O is tuned
- Sufficient CPU for the database
- Sufficient RAM

Database configuration for the hardware
- Correctly size SGA
- Ensure Parallel configuration is set correctly for the database
- Use Data Masking Parallel Execution where possible and enter appropriate Parallel Degree value
- In a RAC environment shutdown all nodes except for one
- Restrict access to database when running masking

Application schema
- Statistics are current on application schema
- Use Guaranteed Restore Point when developing masking algorithms
- Shutdown application and restrict access to database
- Review AWR and ADDM reports prior to running masking
At-Source Data Masking Best Practice

**Application data model (ADM)**
- Make sure ADM has all applications, entities and relationships defined.
- Run Verify on the ADM to ensure consistency between the ADM and the database At-Source Data Masking will run.

**Data masking definition**
- Make sure to have tested the format types used in the masking definition.
- Ensure there is sufficient space for the masked dump file(s).
- Utilize the compression and encryption of the At-Source Data Masking dump file.
Heterogeneous Data Masking Best Practice with Oracle Database Gateways

Oracle Gateway parameter file entries

- Install Oracle Database 11.2.0.3
- Create a new database to stage and mask/subset data from heterogeneous databases
- Install Oracle Gateways for `<heterogeneous db>` v11.2.0.3 in a separate ORACLE_HOME on the same server
- Deploy Enterprise Manager agent to this server
- Configure gateway initialization file found located in: `$<gateway_home>/dg4msql/admin/initdg4sql.ora`
**Heterogeneous Data Masking Best Practice with Oracle Database Gateways**

**Oracle Gateway parameter file entries**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>_HS_FDS_HGOEXEC_BUNDLING=True</code></td>
<td>For 11.2.0.3 and above only and don’t forget the “_” at the beginning of the parameter name</td>
</tr>
<tr>
<td><code>HS_FDS.Fetch.Rows</code></td>
<td>The default is 100. It represents number of rows that the gateway fetch in a batch from the non-oracle database. If you plan to select huge amount of sybase data, increase it to a large amount (e.g., 2000 or more)</td>
</tr>
<tr>
<td><code>HS_RPC.FETCH.SIZE</code></td>
<td>Default is 50000. It represents number of bytes of the buffer that the gateway send to Oracle database as a batch. Increase to a large number (e.g., 500000 or more).</td>
</tr>
</tbody>
</table>
Example 1: At-Source Masking in Snapshot Standby Database

Mask the most current production data without being on prod db

- Oracle introduced At-Source Data Masking in Oracle Enterprise Manager 12c with DB-Plug-in 12.1.0.3
- Allows you to mask production data as it is exported out of production database and saved into data pump export files
- Concern:
  - You would like to have the latest masked production data however you do not want to run At-Source masking on production
- Solution
  - Snapshot standby database is a fully updateable standby database closest to that of production
  - Run At-Source Masking on snapshot standby database
Example 1: At-Source Masking in Snapshot Standby

1. Convert physical standby database to snapshot standby database
2. Execute At-Source data masking (and/or subsetting) on snapshot standby database
3. Convert snapshot standby database to Physical standby and start recovery
4. Import masked and/or subsetted data with non-production users
Example 1: At-Source Masking in Snapshot Standby

1. Convert physical standby database to snapshot standby database
2. Execute At-Source data masking (and/or subsetting) on snapshot standby database
3. Convert snapshot standby database to Physical standby and start recovery
4. Import masked and/or subsetted data with non-production users
Example 1: At-Source Masking in Snapshot Standby

1. Convert physical standby database to snapshot standby database
2. Execute At-Source data masking (and/or subsetting) on snapshot standby database
3. Convert snapshot standby database to Physical standby and start recovery
4. Import masked and/or subsetted data with non-production users
Example 1: At-Source Masking in Snapshot Standby

1. Convert physical standby database to snapshot standby database
2. Execute At-Source data masking (and/or subsetting) on snapshot standby database
3. Convert snapshot standby database to Physical standby and start recovery
4. Import masked and/or subsetted data with non-production users

4- Import masked and/or subsetted data into non-production database
Example 2: Advanced Technique for data masking

User defined function or SQL expression format type

- Data Masking provides an assortment of out of the box format type
- To decide which one to use may result in a choice between two formats
  - For example, to mask a tables columns based on a value in another table, one can accomplish this by using
    - User Defined Function or
    - SQL Expression
- For best practice, if you can use SQL Expression over a UDF, use a SQL Expression. This will have the greatest performance and ease of maintenance
- An example follows…
Example 2: Advanced Technique for data masking

SQL expressions

- Mask data in a table based on a value in a column in another table
- Example: we have three tables:
  - ACTOR contains a list of starring actors in Hollywood
  - ACTOR_ROLE contains the various roles an actor plays
  - ROLE is a master table describing the role an actor plays
- Objective:
  - Mask name in ACTOR based on specific roles in ACTOR_ROLE

```
<table>
<thead>
<tr>
<th>ACTOR_ROLE</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ACTOR_ID</td>
<td>NUMBER(2)</td>
</tr>
<tr>
<td>ROLE_ID</td>
<td>VARCHAR2(2)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ROLE</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ROLE_ID</td>
<td>VARCHAR2(2)</td>
</tr>
<tr>
<td>ROLE_DESC</td>
<td>VARCHAR2(25)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ACTOR</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ACTOR_ID</td>
<td>NUMBER(2)</td>
</tr>
<tr>
<td>ACTOR_NAME</td>
<td>VARCHAR2(35)</td>
</tr>
</tbody>
</table>
```
• Create an Application Data Model
• Add sensitive column ACTOR.ACTOR_NAME
• Create a data masking definition on ADM
• For ACTOR_NAME column add format type of SQL Expression
• Generate and execute masking definition

```
case when %actor_id% in ( 
  select ar.actor_id 
  from OW13.actor_ROLE ar, 
      OW13.ROLE r 
  where ar.actor_id = %actor_id% 
  and ar.role_id = r.role_id 
  and r.role_id in ('PK','TK','FO')) 
then %actor_NAME% 
else dbms_random.string('A',10) 
end

Condition to not mask name if the role id is in PK,TK or FO
```

For others mask with a random string

```
SQL> select * from actor_role;
ACTOR_ID  ROLE
--- ------
1       SK
2       KT
3       SK
4       PK
5       TK
6       TK
8 rows selected.

SQL> select * from actor;
ACTOR_ID  ACTOR_NAME
---------  -----------------
1       Maureen OHara
2       Kirk Douglas
3       Dick Van Dyke
4       Shirley Temple
5       James Garner
6       Sean Connery
6 rows selected.
```

Advanced use of SQL Expression Format Type
New Features
New Features

Data Masking definition
- Ability to filter columns in a definitions. This allows to quickly search for masking columns in the definition
- Enable/Disable columns in masking definition. A masking definition can contain a superset of columns to be masked however when generating the script, enabled columns will only be included in the masking process

Expanded capability of masking CLOBs/BLOBS
- This is done by defining regular expression to search for sensitive data and mask. Multiple values can be searched of in a column and respectively masked

Adaptive Referential Sizing
- Ability to visualize data selection criteria in subset definition
- Greater accuracy in estimating subset sizing

Test Data Management
- Ability to visualize data selection criteria in subset definition
- Greater accuracy in estimating subset sizing

Command line Support of TDM
- Execute subset definitions and other administrative tasks using emcli verbs

Masking and subsetting Templates
- Templates are made available in Oracle Enterprise Manager through Self-Update
- This method allows for out-of-band template releases from Enterprise Manager release

Self-Update
- Templates are made available in Oracle Enterprise Manager through Self-Update
Data Masking Template

- E-Business Suite 12.1: Available on Oracle Enterprise Manager 11g and 12c
- Oracle Health Insurance Back Office: Available on Oracle Enterprise Manager 12c
- PeopleSoft: To be released in PeopleSoft 9.2
Customer Case Study
Enterprise Data Masking Solution

Tom Jernejcic
Supervising Database Administrator, RCIT
• Where is Riverside County?
• Technical environment
• Business driver
• Options available
• Business Case
• Implementation
• Testing
• Rollout
• Lessons learned
• Benefits
Where is Riverside County?

Who are we ...

- Fifty miles east of Los Angeles
- 7.2 million square miles
- Population of 2.2 million
- Fourth Largest County in California
- Diverse economy
  - Agriculture
  - Manufacturing
  - Distribution
  - Education
Technical Environment

ERP implementation…

- PeopleSoft Financials and HRMS/Payroll
- IBM P570 (P6/11 core), AIX 5.3/6.1, VMware, MS Windows
- Oracle 11gR2 11.2.0.3
- Enterprise Manager 12c
- 43 test & development databases
- 18 production databases
- Single and multi node RAC clusters
- Data Warehousing/Reporting
The challenge at hand...

- Limiting exposure to PII data
  - Solicited
  - Unsolicited
- Protecting propagated PII data
  - From
    - Consultants
    - Employees
  - Functional testing
  - Interface testing
Options Available

Choices...

• Create custom data masking modules
  – Cumbersome
  – Difficult to maintain
  – Schema changes may require code modifications
  – Limited resources and support
• Implement third party (non-Oracle) solution
• Implement Oracle Data Masking
Why Oracle data masking ...

- Eliminates most/all custom processes
- Seamlessly lives within OEM
  - Common interface
  - Maintained by Oracle
  - Supportability
    - OEM upgrades in step with DB upgrades
  - Minimal training
- Compliments other Oracle security offerings (e.g. ASO)
- Delivered masking definitions
- Customizable where necessary
- Easy to accommodate schema changes
- Addresses referential integrity requirements
- Simplified execution
- Effortless integration with existing cloning processes
Implementation

Steps...

• Identify sensitive data and mask requirements
  – Review results and constraints with ISO
• Deploy EM delivered Data Management packages
• Create Application Data Model (ADM)
• Configure
  – Masking definition
  – Masking format
• Generate masking scripts
Testing

Steps...

• Clone production to test
• Back up test database
• Run initial tests without masking
  – Focus on columns to be masked
• Restore test database
• Execute data masking scripts
• Rerun same tests
• Compare results
Rollout

Steps...

• Create generalized master script
  – Environment independent

• Update clone process
  – Add call to data masking master script

• Modify clone request form
  – Include option to mask or not mask*
  * Special authorization required
What did we learn...

• Validate masking requirements against application requirements and constraints

• “to mask or not to mask”
  – Leave mask generation process optional

• Any DDL changes to masked columns will require a script regeneration

• Data size and number of columns masked will influence cloning run times
Benefits

What did we gain...

• Sensitive public and employee data protected from internal and external attacks

<table>
<thead>
<tr>
<th>Environment</th>
<th># Tables</th>
<th># Columns</th>
<th># Rows</th>
<th>Elapsed Time (hrs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Financials</td>
<td>41</td>
<td>69</td>
<td>26,035,655</td>
<td>4</td>
</tr>
<tr>
<td>HRMS</td>
<td>46</td>
<td>251</td>
<td>35,141,931</td>
<td>2</td>
</tr>
</tbody>
</table>

• Custom complex masking techniques eliminated reducing time and resources

• Automation reduces errors and insures masking processes are executed successfully
Any Questions?
## Database Manageability

### Recommended Sessions

<table>
<thead>
<tr>
<th>Session</th>
<th>General Session</th>
<th>Day</th>
<th>Time</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>GEN8792</td>
<td>General Session: Database Management Innovations - Oracle Database 12c Manageability Highlights</td>
<td>Wednesday</td>
<td>10:15 AM</td>
<td>Moscone South – 103</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Session</th>
<th>Session</th>
<th>Day</th>
<th>Time</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>CON9582</td>
<td>Oracle Exadata Management Deep Dive with Oracle Enterprise Manager 12c</td>
<td>Monday</td>
<td>12:15 PM</td>
<td>Westin - Metropolitan I</td>
</tr>
<tr>
<td>CON9573</td>
<td>Managing the Oracle Identity Management Platform with Oracle Enterprise Manager</td>
<td>Monday</td>
<td>1:45 PM</td>
<td>Moscone South - 130</td>
</tr>
<tr>
<td>CON8788</td>
<td>Maximizing Database Performance with Database Replay</td>
<td>Tuesday</td>
<td>10:30 AM</td>
<td>Moscone South - 308</td>
</tr>
<tr>
<td>CON9583</td>
<td>Harness the Power of Oracle Database 12c with Oracle Enterprise Manager Database as a Service</td>
<td>Tuesday</td>
<td>3:45 PM</td>
<td>Moscone South – 305</td>
</tr>
<tr>
<td>CON9579</td>
<td>Step-by-Step Cookbook for Identifying and Tuning SQL Problems</td>
<td>Wednesday</td>
<td>1:15 PM</td>
<td>Moscone South – 103</td>
</tr>
<tr>
<td>CON4666</td>
<td>Oracle Enterprise Manager 12c Database Lifecycle Management Automatic Provisioning and Patching</td>
<td>Wednesday</td>
<td>3:30 PM</td>
<td>Moscone South – 131</td>
</tr>
<tr>
<td>CON8768</td>
<td>DBA Best Practices for Protecting Data Privacy with Oracle’s Data Masking</td>
<td>Wednesday</td>
<td>3:30 PM</td>
<td>Moscone West – 2024</td>
</tr>
<tr>
<td>CON9577</td>
<td>Active Session History Deep Dive: Advanced Performance Analysis Tips</td>
<td>Wednesday</td>
<td>3:30 PM</td>
<td>Moscone South – 104</td>
</tr>
<tr>
<td>CON3255</td>
<td>Being Sure: Confident Consolidations with Oracle Real Application Testing 12c</td>
<td>Wednesday</td>
<td>5:00 PM</td>
<td>Moscone South – 306</td>
</tr>
<tr>
<td>CON3103</td>
<td>Real Oracle Real Application Testing: What to Expect and Prepare For</td>
<td>Thursday</td>
<td>12:30 PM</td>
<td>Moscone South - 104</td>
</tr>
</tbody>
</table>
Enterprise Manager Resources

- **Oracle.com**: [http://www.oracle.com/enterprisemanager](http://www.oracle.com/enterprisemanager)
  - Enterprise Manager Customer References
- **OTN**: [http://www.oracle.com/technetwork/oem](http://www.oracle.com/technetwork/oem)
  - Demos on Demand
  - Enterprise Manager Forums

**Enterprise Manager Training from Oracle University**
- Oracle Enterprise Manager 12c Learning Library
- Oracle Enterprise Manager Cloud Control 12c: Install and Upgrade
- Using Oracle Enterprise Manager Cloud Control 12c Ed 1
- Oracle Enterprise Manager 12c: Overview Bundle Self-Study
- Oracle Enterprise Manager 12c: Management Bundle Self-Study

**Partners**
- List of Enterprise Manager Specialized Partners
- Oracle Enterprise Manager OPN Knowledge Zone
- Oracle Enterprise Manager OPN Specialization
- Oracle Enterprise Manager Application Quality Management OPN Specialization
- IOUG Oracle Enterprise Manager Special Interest Group

**Social Media**
- Twitter, Facebook, YouTube, Linkedin, Blog