Data Masking: The Ultimate DBA Survival Tool in the Modern World

Jagan R. Athreya
Oracle Corporation

Ravi Meda
Qualcomm, Inc.
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.
Oracle Enterprise Manager
Top-Down, Integrated Application Management

• Complete, Open, Integrated Management for Oracle Technologies
  – Deep, Optimized, Best of Breed
  – Database, Middleware, Packaged Applications, Physical and Virtual Infrastructure

• Business Centric, Top Down Application Management

• Complete Lifecycle Management

• Scalable Grid and Cloud Management
  – Manage many as one
Agenda

• Cost of Data Privacy Breaches
• Implementing Oracle Data Masking
• Customer Case studies
More data than ever…

Source: IDC, 2008
More breaches then ever…

Once exposed, the data is out there – the bell can’t be un-rung

PUBLICLY REPORTED DATA BREACHES

630% Increase

Source: DataLossDB, 2009

Oracle Confidential
More threats than ever…
More Regulations Than Ever…


90% Companies behind in compliance
• 89% of companies use production customer data - often exceeding 10M records - for testing, development, support, training, etc.

• 74% use consumer data, 24% use credit card numbers!!

• Only 23% do anything to suppress sensitive information and 81% relied on contractual clauses to protect live data transferred to outsourcers and other third parties

• 23% said live data used for development or testing had been lost or stolen and 50% had no way of knowing
Business Drivers for Data Masking

- Social Security numbers
- Customer credit cards
- Employee salaries
- Bank a/c numbers
- Corporate accounting
- Patient health data

- Application Developers
- 3rd party IT Service Providers
- Business partners
- Market Research
- Clinical Research

- Sarbanes Oxley
- PCI-DSS
- HIPAA
- GLBA
- California Data Security Breach
- EU Data Protection Directive
What is Data Masking?

**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 non-production environments when the data is shared with non-production users without revealing sensitive information.

<table>
<thead>
<tr>
<th>LAST_NAME</th>
<th>SSN</th>
<th>SALARY</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGUILAR</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>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LAST_NAME</th>
<th>SSN</th>
<th>SALARY</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANSKEKSL</td>
<td>111-23-1111</td>
<td>60,000</td>
</tr>
<tr>
<td>BKJHHEIEDK</td>
<td>222-34-1345</td>
<td>40,000</td>
</tr>
</tbody>
</table>
Agenda

- Cost of Data Privacy Breaches
- Implementing Oracle Data Masking
- Customer Case studies
Data Masking Methodology

1. Find

2. Assess

3. Secure

4. Test
Data Masking Methodology

1. **Find**
Find and Catalog Sensitive Data

Data Finder tool

1. Data Finder Patterns
   - Define pattern match rules for Tables, columns and data
   
   **Table Name:** "EMP*"
   **Column Name:** "*SSN*"
   **Data Format:** ### - ## - ####

2. Enterprise Data Sources
   - Search against selected Oracle Databases

3. Data Finder Reports
   - Results rendered by confidence factor
   - Relevant database fields imported into the Data Privacy Catalog

4. Data Privacy Catalog
   - New database fields added and then protected

- PERSON_SSN, EMP_SSN, SOC_SEC_NUM
Data Masking Methodology

2. Assess
Define Mask Formats and Register in Library

- **Mask Format Library**
  - Mask formats for commonly masked data such as Credit Card number, Social Security Numbers, etc.

- **Mask Primitives to extend Format Library**
  - Random Number
  - Random String
  - Random Date within range
  - Shuffle
  - Sub string of original value
  - Table Column
Leverage 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.Dvrie@spambob.com
- Daniel.118074087.Bates@spambob.com
- Dieter.421172061.Elliott@spambob.com
- Mani.268599080.Field@spambob.com
- Gena.417785060.Altman@spambob.com
Extend with Sophisticated Masking Techniques

• Compound Masks
  – Sets of related columns masked together e.g. Address, City, State, Zip, Phone

• Condition-based Masking
  – Specify separate mask format for each condition, e.g. driver’s license format for each state

• Deterministic Masking
  – Consistent repeatable masking e.g. John always masks to Joe across multiple databases
Ensure Referential Integrity for the Data

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>
Data Masking Methodology

3. Secure
## Separate Duties between App Admin and DBA

<table>
<thead>
<tr>
<th>App Admin</th>
<th>DBA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Identify Sensitive Information</strong></td>
<td><strong>Clone Prod to Staging</strong></td>
</tr>
<tr>
<td><strong>Associate mask format with sensitive information</strong></td>
<td><strong>Execute Mask</strong></td>
</tr>
<tr>
<td><strong>Format Library</strong></td>
<td><strong>Mask Definition</strong></td>
</tr>
</tbody>
</table>
Integrate with Data Center Processes

- Secure Clone-and-Mask workflow
  - Integrated process to create test databases from production
  - After cloning DB in RESTRICTED mode till masking complete

- Privilege Delegation Support
  - Allows mask execution using sudo or PowerBroker

- Masking script directory specification
  - Allows DBAs to specify directory location when masking script should be generated
Data Masking Methodology

1. Find
2. Assess
3. Secure
4. Test
Customize Mask and Test

- **Post-Mask SQL**
  - for LOBs, attachments, summary values

- **Comparing before & after values**
  - To save the mapping tables to compare before and after values after a mask run during testing

- **REDO log generation**
  - To allow FLASHBACK to pre-masked state when testing masking routines.
Masking Process – Internals

1. Capture and disable Constraints on “sensitive” table
2. Build mapping table containing original sensitive and masked values using masking routines
3. Rename “sensitive” table
4. Recreate masked table from original table replacing sensitive with masked values from mapping tables using CTAS
5. Collect statistics
6. Restore constraints based on original table
7. Drop Renamed table and mapping table
8. Capture and disable Constraints on “sensitive” table

ORACLE
High Performance Execution

- Linux x86 4 CPU: Single core Pentium 4 (Northwood) [D1])
- Memory: 5.7G
- Column scalability
  - 215 columns masked across 100 tables
  - 60GB Database
  - 20 minutes
- Rows scalability
  - 100 million row table, 6 columns masked
  - Random Number
  - 1.3 hours
Specify Execution Options

• Statistics Refresh
  – To enable DBAs to run their own custom statistics generation routine

• Degree of Parallelism
  – To optimize the performance of the mask execution based on the number of processors available
Validate Mask and Generate Script

- Ensure uniqueness can be maintained
- Ensure formats match column data types
- Check Space availability
- Warn about Check Constraints
- Check presence of default Partitions
- Generate PL/SQL-based masking script upon successful validation
Data Masking Implementation Continuum

- Privacy Catalog
- Application Discovery
- Mask Development
- Test System Automation
- Application Testing

- Privacy Catalog
- Application Discovery
- Mask Development
- Test System automation
- Application Testing

- Privacy Catalog
- Mask Templates
- Mask Development
- Test System Automation
- Application Testing

- Privacy Catalog
- Mask Templates
- Mask Development
- Test System Automation
- Application Testing
Agenda

- Cost of Data Privacy Breaches
- Implementing Oracle Data Masking
- Customer Case study
**UK-based Government Agency**

Data Masking Pack delivered rapid compliance of non-production eBusiness Suite environments

<table>
<thead>
<tr>
<th>Business Challenges</th>
<th>Solution</th>
<th>Business Results</th>
</tr>
</thead>
</table>
| • Internal audit assessment indicated non-compliance with established privacy standards  
  • Personnel information at risk of being exposed to non-production users  
  • Needed to bring all their Oracle eBusiness Suite non-production environments compliant within short remediation period to pass the audit | • Data Masking Pack provided flexible routines to mask various types of sensitive data  
  • IT team leveraged the extensibility to add user-define masking routines to meet their needs | • Successfully met the audit requirements within 4 weeks of identifying non-compliance  
  • Enabled personnel data in eBusiness Suite application to be shared with non-production users in line with established standards  
  • Provided a successful proof-point for masking Oracle eBusiness Suite applications |
EMEA-based Real Estate Company
Data Masking Pack accelerated availability of production data for testing while improving DBA productivity

<table>
<thead>
<tr>
<th>Business Challenges</th>
<th>Solution</th>
<th>Business Results</th>
</tr>
</thead>
</table>
| • Custom scripts to mask sensitive data were not able to scale to meet growing data volumes  
• DBA team under increasing pressure to make production data available to for application testing within short time frames | • Data Masking Pack delivered an out-of-the-box solution to replace custom database scripts  
• High performance masking capabilities accelerated masking process from 6 hours using database scripts to 6 minutes using Data Masking Pack | • 60 X performance improvement in masking process resulted in faster turnaround of test system creation  
• Improved DBA productivity by eliminated the requirement to maintain custom scripts |
Oracle Data Masking Solution Using Oracle Enterprise Manager

Ravi Meda
Qualcomm Inc.
Agenda

1. Overview of OEM Grid control Infrastructure
2. Current Data Scrambling issues
3. Oracle Data Masking Implementation
4. Best Practices and Benefits
Grid Control Setup

Database Servers:
- Database Server (Solaris)
- Database Server (Linux)
- Database Server (Solaris)
- Database Server (Solaris)
- Database Server (Linux)

Load Balancer: http://oemgrid-prod/em

Application Servers in
Active/Active Cluster
OMS: 10.2.0.5.0

Oracle RAC Database
RDBMS: 10.2.0.4.0

Agents: 10.2.0.5.0
Overview of OEM Grid control Infrastructure

- Currently using 10gR5 OMS
- OMS is an active-active cluster on Linux Hardware
- Repository database is on 10.2.0.4 with RAC
- Hundreds of targets were configured in OMS
- Dedicated OMS for Prod databases and Non-Prod Databases.
1. Overview of OEM Grid control Infrastructure

2. Current Data Scrambling issues

3. Oracle Data Masking Implementation

4. Best Practices and Benefits
Data is sent offshore for application testing.
Scrambling is done via custom scripts after refresh.
Developers who wrote the scripts had access to production data before scrambling.
Current Data Scrambling issues

- Manual scripts run by developer
- Not 100% compliant with industry
- No referential integrity is maintained
- Data issues
1. Overview of OEM Grid control Infrastructure

2. Current Data Scrambling issues

3. Oracle Data Masking Implementation

4. Best Practices and Benefits
Database Services

Masking Implementation – Privacy Attributes

Employee
Personal data
Dependent Benefits
Employment details
Non-employee workforce
Recruitment candidate
Temporary workforce
Relocation
Tables with Sensitive Data: 98 to 120 columns

Masking Formats:
- Employee related
- Non-employee workforce

Database size: $\frac{1}{2}$ TeraByte

Custom database privileges granted for masking

Masking job execution: 30-40 minutes
Separation of duties

- HR analyst defines the mask definition
- Operator submits the job to clone Production to Test and mask.
- DBA monitors the execution

Easy to use and works great for referential integrity

Automatic alerts when

- insufficient space in SYSTEM or TEMP or data
- not enough privileges to do masking

Custom data masking script now RETIRED.
1. Overview of OEM Grid control Infrastructure

2. Current Data Scrambling issues

3. Oracle Data Masking Implementation

4. Best Practices and Benefits
Best Practices and Benefits

- Leverage format libraries to store data masking definitions
- All the scrambled data is 100% compliant
- Re-run the failed job
- Still have the old data in the table for verification.
Oracle Database Security
Defense-in-Depth for Security and Compliance

Monitoring
Configuration Management
Audit Vault
Total Recall

Access Control
Database Vault
Secret
Label Security

Encryption and Masking
Advanced Security
Secure Backup
Data Masking
<table>
<thead>
<tr>
<th>Company</th>
<th>Benefit</th>
</tr>
</thead>
<tbody>
<tr>
<td>NetApp</td>
<td>Deploys SOA infrastructure 92% faster</td>
</tr>
<tr>
<td>Nokia</td>
<td>Saves 80% time and effort for managing Databases</td>
</tr>
<tr>
<td>TomTom</td>
<td>Avoids online revenue losses up to 25%</td>
</tr>
<tr>
<td>AmTrust Bank</td>
<td>Improves IT productivity by 25%</td>
</tr>
<tr>
<td>Commonwealth Bank</td>
<td>Drives asset utilization up by 70%</td>
</tr>
<tr>
<td>Cummins</td>
<td>Cuts configuration management effort by 90%</td>
</tr>
<tr>
<td>Telstra</td>
<td>Saves $1.9 million with Oracle Enterprise Manager</td>
</tr>
<tr>
<td>SSM</td>
<td>Saves $170,000 per year with Oracle Enterprise Manager</td>
</tr>
<tr>
<td>City University</td>
<td>Replaces manual tools with automation; saves time by 50%</td>
</tr>
<tr>
<td>STARWOOD</td>
<td>Reduces Database testing time by 90%</td>
</tr>
<tr>
<td>Bayer</td>
<td>Reduces provisioning effort by 75%</td>
</tr>
<tr>
<td>Meridian</td>
<td>Saves weeks on application testing time</td>
</tr>
<tr>
<td>CUMIS</td>
<td>Cuts application testing from weeks to hours</td>
</tr>
<tr>
<td>APL</td>
<td>Reduces critical patching time by 80%</td>
</tr>
<tr>
<td>City of Evanston</td>
<td>Delivers 24/7 uptime with Oracle Enterprise Manager</td>
</tr>
</tbody>
</table>