Oracle Data Masking and Subsetting
Minimize Risk, Maximize Value
Agenda

1. Data, Data Proliferation, and Security Challenges
2. Oracle Data Masking and Subsetting Capabilities
3. Functional Highlights
4. Masking and Subsetting for Cloud
Data is today’s capital
“The world’s most valuable resource is no longer oil, but data”

Data drives
• Analytics and automation
• Advertising and marketing budgets
• Personalization and improved experience
• Business analytics and decisions
• Government policies and plans

Overall, data helps improve products and services, provide better user experience, and support and grow businesses
Data can be a liability
The scary side of data economy

Data breaches are exploding worldwide

- Database is the most common asset involved in breaches

Data losses can be catastrophic for businesses impacting

- Finances due to compensations, penalties, legal, PR, share value, recovery cost
- Brand reputation, customer trust, intellectual property, competitiveness
- Overall business and revenue

Fast evolving, stringent regulatory landscape

- Across industries and regions
- Laws that aim to protect data and citizen privacy
Proliferation of sensitive data increases security risk
Compliance regulations mandate sensitive data protection

**PCI DSS 3.2.1 May 2018**

6.4.3 Production data (live PANs) are not be used for testing or development purposes.

**EU General Data Protection Regulation**

The application of pseudonymisation to personal data can reduce the risks for the data subjects concerned and help controllers and processors meet their data protection obligations.

**The Health Insurance Portability and Accountability Act of 1996**

The removal of specified identifiers of the individual and of the individual’s relatives, household members, and employers is required.

... and more
Your dilemma
To do, or not to do

Your wish
• Get actionable insights from your data to take smarter business decisions
• Use realistic data for development and analysis
• Quickly share data with developers, data scientists, and partners

Your concern
• Avoid proliferation of sensitive data to non-production environments
• Comply with data privacy regulations/laws such as GDPR
• Minimize time and storage costs

The Solution?
Agenda

1. Data, Data Proliferation, and Security Challenges
2. Oracle Data Masking and Subsetting Capabilities
3. Functional Highlights
4. Masking and Subsetting for Cloud
Oracle Data Masking and Subsetting

Minimize proliferation of sensitive data to non-production environment

Production

SSN
463-62-9832
576-40-7056
518-12-6157
281-50-3106

Credit Card
3715 4691 3277 8399
5136 6247 3878 3201
3599 4570 2897 4452
9331 3219 2331 9437

Non-Production

SSN
463-62-9832
555-12-1234

Credit Card
3715 4691 3277 8399
5555-5555-5555-4444

Sensitive Data Discovery
Comprehensive Masking Options
Goal/Condition Based Subsetting
In-Database or In-Export Masking
Support for Cloud and Non-Oracle DBs
Workload Capture & Clone Masking
Pre-installed in Enterprise Manager
Sensitive Data Discovery
Understand what’s at risk

Identification
- SSN
- Name
- Email
- Phone
- Passport
- DL
- Tax ID
- ...

Biographic
- Age
- Gender
- Race
- Citizenship
- Hostname
- Address
- Family Data
- Date of Birth
- Place of Birth
- ...

IT
- IP Address
- User ID
- Password
- Hostname
- GPS location
- ...

Financial
- Credit Card
- User ID
- CC Security PIN
- Bank Name
- Bank Account
- IBAN
- Swift Code
- ...

Healthcare
- Provider
- Insurance
- Height
- Blood Type
- Disability
- Pregnancy
- Test Results
- ICD Code
- ...

Employment
- Employee ID
- Job Title
- Department
- Hire Date
- Income
- Stock
- ...

Academic
- College Name
- Grade
- Student ID
- Financial Aid
- Admission Date
- Graduation Date
- Attendance
- ...
Application Data Modeling
Sensitive Data Discovery

Database Defined Relations

Application Defined Relations

Consumed By

Subsetting
Encryption
Masking
Database Vault
Redaction
Auditing

Automated Discovery

Data Governance

Oracle Enterprise Manager

Data Relationships
Sensitive Columns

Data Governance
Data Masking
Minimize risk by masking sensitive data

Replaces sensitive data with fictitious, but realistic, data that remains useful for analytics, development, and other purposes

Preserves data integrity, helping ensure that apps continue to work with the masked data

Provides comprehensive and flexible masking options to meet diverse requirements

Masking templates for selected Oracle applications
Data Masking
Comprehensive and flexible masking formats

Provides common predefined masking formats
  • Credit Card Number
  • Social Security Number
  • National Insurance Number
  ... and more

Provides flexibility to customize masking formats
  • Fixed number / string
  • Random numbers / strings / dates / list
  • Substitute, Encrypt, Shuffle, Nullify
  • User Defined PL/SQL Function
  ... and more

Sample masked values help you visualize and validate the masked data
## Data Masking

Masking transformations to meet diverse business use cases

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Conditional masking</strong></td>
<td>Masks rows differently based on condition</td>
<td>Mask national identifiers based on country</td>
</tr>
<tr>
<td><strong>Deterministic masking</strong></td>
<td>Masks data to the same consistent values across multiple databases or masking jobs</td>
<td>Mask employee identifiers consistently across schemas and databases</td>
</tr>
<tr>
<td><strong>Compound masking</strong></td>
<td>Ensures masked values across related columns retain the same relationship</td>
<td>Mask address fields such as state, postal code, and country as a group</td>
</tr>
<tr>
<td><strong>Format preserving</strong></td>
<td>Masks data while preserving its format such as length and special characters</td>
<td>Mask tax identifiers while preserving spaces and hyphens</td>
</tr>
<tr>
<td><strong>Reversible masking</strong></td>
<td>Encrypts and decrypts data using cryptographic key</td>
<td>Unmask data after receiving the processed data from a partner</td>
</tr>
<tr>
<td><strong>Shuffling</strong></td>
<td>Shuffles the values within a column</td>
<td>Shuffle age of employees in an organization</td>
</tr>
<tr>
<td><strong>Perturbation</strong></td>
<td>Generates random values within a user-provided range</td>
<td>Generate random dates within a specified data range</td>
</tr>
</tbody>
</table>
Data Masking

Examples

**Mask based on conditions**

<table>
<thead>
<tr>
<th>Country</th>
<th>Identifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>CA</td>
<td>226-956-324</td>
</tr>
<tr>
<td>US</td>
<td>610-02-9191</td>
</tr>
<tr>
<td>UK</td>
<td>JX 75 67 44 C</td>
</tr>
</tbody>
</table>

**Shuffle records**

<table>
<thead>
<tr>
<th>Health Records</th>
<th>Health Records</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Generate deterministic output**

<table>
<thead>
<tr>
<th>HR</th>
<th>Emp ID</th>
<th>First Name</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>324</td>
<td>Charlie</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>FIN</th>
<th>Emp ID</th>
<th>First Name</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>324</td>
<td>Murali</td>
</tr>
</tbody>
</table>

**Generate random values while preserving format**

<table>
<thead>
<tr>
<th>Name</th>
<th>License#</th>
</tr>
</thead>
<tbody>
<tr>
<td>Richard</td>
<td>7ZPN788</td>
</tr>
<tr>
<td>Rishabh</td>
<td>DL 12TC 0204</td>
</tr>
</tbody>
</table>

**Mask operating system files stored as LOBs**

<table>
<thead>
<tr>
<th>LOB</th>
<th>Search</th>
<th>Replace</th>
</tr>
</thead>
<tbody>
<tr>
<td>3178973456</td>
<td>[0-9]{10}</td>
<td>**</td>
</tr>
<tr>
<td>6509876745</td>
<td></td>
<td>**</td>
</tr>
</tbody>
</table>
Data Masking
Masking templates

Oracle E-Business Suite Templates

• Available for specific versions
• Covers multiple modules
  - HCM, FIN, ATG, MOC, MSC, OKL, OPM, PRC, PROJ, and TCA
  - Around 1000 columns
• Masks key PII elements
  - Name, Location, Phone, Accounts, Drivers License, National ID, Website, Compensation, Nationality, Health Information, Audit information, and more ...
• Includes predefined data model and masking rules
• Included in product license

EBS Masking MOS Note 2076834.1
Data Subsetting
Why subset data?

- Minimize risk by sharing only relevant data with internal and external teams
- Reduce storage cost in non-production environments
- Extract subscriber data from SaaS
- Perform research and analysis on a subset of data
- Extract subset of data as part of e-discovery requests
Data Subsetting
Goal or condition based subsetting

Relative Table Size
- 100M Rows
- 20M Rows
- 2M Rows

Condition Based
Extract ASIA Sales

Table Partitions
JAN
SALES
EMEA
APAC
FEB

SALES
JAN
FEB
# Data Subsetting

Preview and validate results

<table>
<thead>
<tr>
<th>Name</th>
<th>Table Rule</th>
<th>Source Size</th>
<th>Estimated Subset Size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>MB</td>
<td>Rows</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rows</td>
<td>MB</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rows</td>
<td>%</td>
</tr>
<tr>
<td>Applications and Tables</td>
<td></td>
<td>912.1908</td>
<td>8661245</td>
</tr>
<tr>
<td>TDM(TDM)</td>
<td></td>
<td>912.1908</td>
<td>8661245</td>
</tr>
<tr>
<td>TDM(TDM)</td>
<td></td>
<td>483.5967</td>
<td>443883</td>
</tr>
<tr>
<td>TDM(TDM)</td>
<td></td>
<td>53.01</td>
<td></td>
</tr>
<tr>
<td>H_LINEITEM</td>
<td></td>
<td>606.6597</td>
<td>6001215</td>
</tr>
<tr>
<td>H_ORDER</td>
<td>o_custkey in (select c_custkey from tdm.h_custo...</td>
<td>148.7732</td>
<td>1500000</td>
</tr>
<tr>
<td>H_PARTSUPP</td>
<td></td>
<td>109.1003</td>
<td>800000</td>
</tr>
<tr>
<td>H_PARTSUPP</td>
<td></td>
<td>109.1003</td>
<td>800000</td>
</tr>
<tr>
<td>H_CUSTORMER</td>
<td></td>
<td>16.1358</td>
<td>100712</td>
</tr>
<tr>
<td>H_CUSTORMER</td>
<td></td>
<td>67.14</td>
<td></td>
</tr>
<tr>
<td>H_PART</td>
<td></td>
<td>22.316</td>
<td>200000</td>
</tr>
<tr>
<td>H_SUPPLIER</td>
<td></td>
<td>22.316</td>
<td>200000</td>
</tr>
<tr>
<td>H_SUPPLIER</td>
<td></td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>H_NATION</td>
<td></td>
<td>0.0021</td>
<td>25</td>
</tr>
<tr>
<td>H_NATION</td>
<td></td>
<td>0.0021</td>
<td>25</td>
</tr>
<tr>
<td>H_REGION</td>
<td></td>
<td>0.0004</td>
<td>5</td>
</tr>
<tr>
<td>H_REGION</td>
<td></td>
<td>0.0004</td>
<td>5</td>
</tr>
<tr>
<td>H_REGION</td>
<td></td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>
Agenda

1. Data, Data Proliferation, and Security Challenges
2. Oracle Data Masking and Subsetting Capabilities
3. Functional Highlights
4. Masking and Subsetting for Cloud
Streamlined Management

Preinstalled with Oracle Enterprise Manager

Centralized and unified management
- Unified interface for sensitive data discovery, masking, and subsetting
- Scheduling and reporting

Non-Oracle databases supported through Oracle Database Gateways
Flexible Deployment Options

**In-Database**
- Clone → Mask/Subset → Replicate

- Source
- Staging
- Target

**In-Export**
- Clone → Mask/Subset → Replicate

- Source
- Export File
- Target

Minimal impact on the production environment

Sensitive data remains within the production perimeter
## Data Masking Integrations

<table>
<thead>
<tr>
<th>Integration Point</th>
<th>Benefit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Database Lifecycle Management Pack</td>
<td>Integrated clone and mask</td>
</tr>
<tr>
<td>Cloud Management Pack</td>
<td>Integrated snap-clone and mask</td>
</tr>
<tr>
<td>Real Application Testing Pack</td>
<td>Mask in database workload captures</td>
</tr>
<tr>
<td>Oracle Data Integrator</td>
<td>Mask incrementally captured data</td>
</tr>
</tbody>
</table>
Implementation Methodology

- Discover Sensitive Data
- Discover Data Relationships

Create Data Model

Select Formats & Criteria
- Masking formats and templates
- Goals & conditions for Subsetting

Preview & Validate
- Preview masking algorithm results
- Preview subset reduction results

Execute Transformation
- In-database or In-Export
- Rerun with same data model
Agenda

1. Data, Data Proliferation, and Security Challenges
2. Oracle Data Masking and Subsetting Capabilities
3. Functional Highlights
4. Masking and Subsetting for Cloud
Mask on Premises and Upload to the Cloud
Oracle Database Cloud Service (PaaS)

Clone → Mask/Subset → Upload

Extract → Mask/Subset → Upload

On Premises

On Premises
Mask on the Wire or in the Cloud
Oracle Database Cloud Service (PaaS)

Clone & Mask PDB to Cloud

Mask/Subset in the Cloud

On Premises
Masking in Oracle Fusion Cloud Service

1. Customer requests to mask sensitive data in dev/test

Oracle Managed Cloud Services team masks the sensitive data using automated procedures
Oracle Data Masking and Subsetting Summary

Key Benefits
- Maximizes business value of data without compromising sensitive data
- Minimizes security risk by not proliferating sensitive data
- Reduces compliance boundary by masking sensitive data
- Reduces storage cost for test/dev by subsetting

Key Features
- Automates sensitive data discovery and maintains referential integrity
- Provides comprehensive and flexible masking options
- Supports goal and condition based subsetting
- Integrates with enterprise data management tools
Thank You
Our mission is to help people see data in new ways, discover insights, unlock endless possibilities.