
ORACLE®

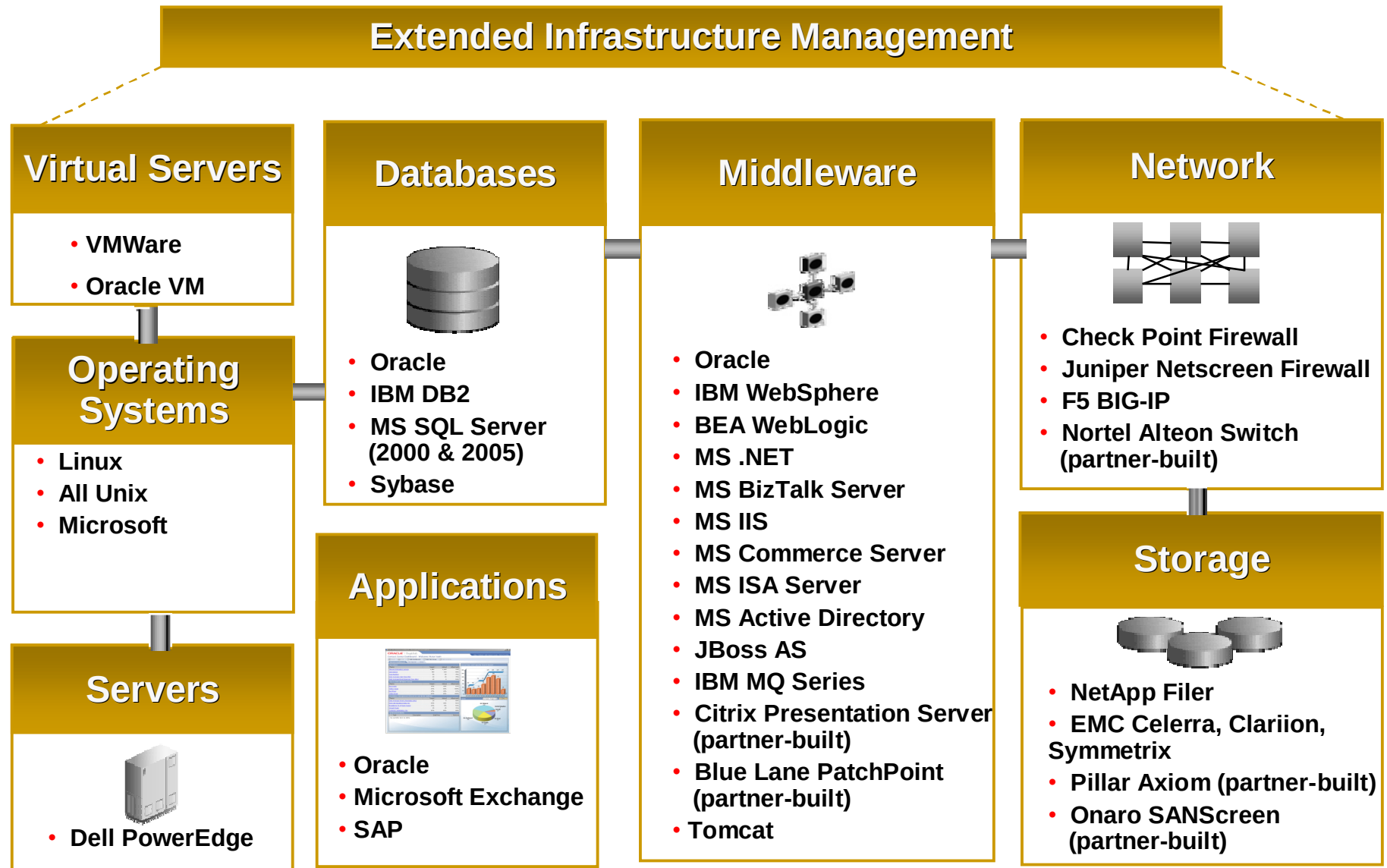


Protecting Privacy with Data Masking Oracle Enterprise Manager 10gR4

Nicholas J. Donatone
Principal Grid Sales Consultant

Grid Control - Infrastructure Management

Broad management for all application-dependent components





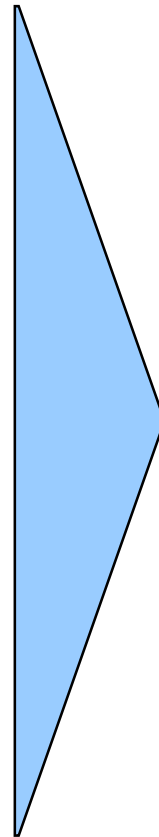
Business Drivers for Data sharing

Application Testing

- Offshore or in-house application development
- Offshore or in-house software QA

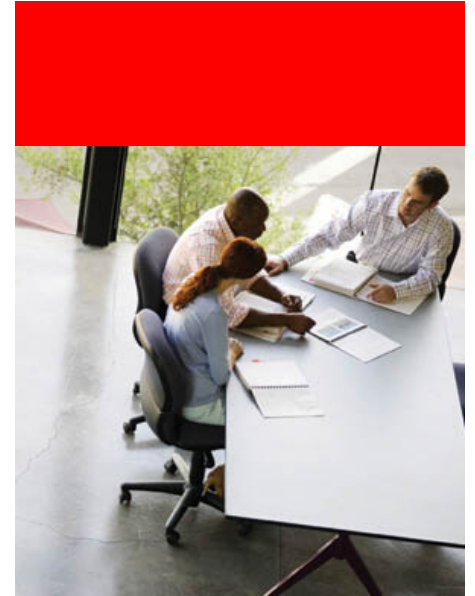
Data sharing

- 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

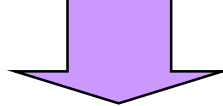
What

- 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.

LAST_NAME	SSN	SALARY
AGUILAR	203-33-3234	40,000
BENSON	323-22-2943	60,000
D' SOUZA	989-22-2403	80,000
FIORANO	093-44-3823	45,000

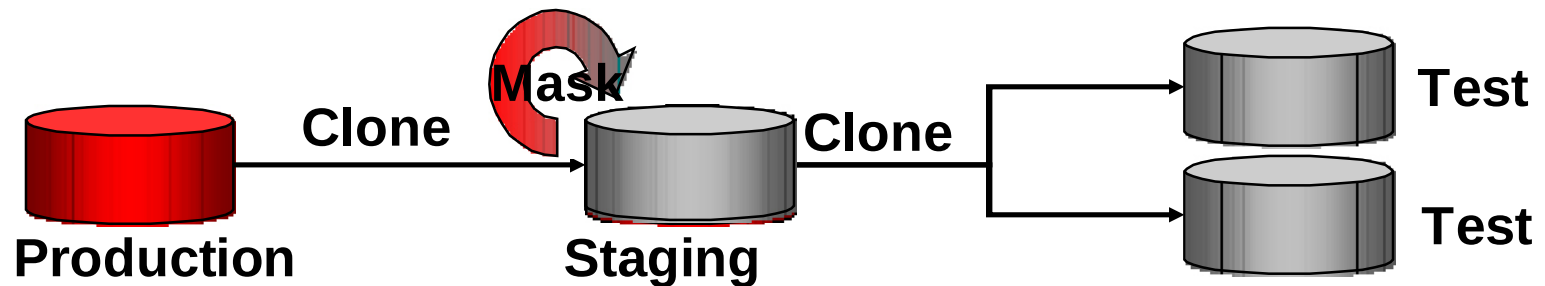
Why

- 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



LAST_NAME	SSN	SALARY
ANSKEKSL	111-23-1111	40,000
BKJHHEIEDK	111-34-1345	60,000
KDDEHLHESA	111-97-2749	80,000
FPENZXIEK	111-49-3849	45,000

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

Masking Definition: Define Format

Database: database Logged In As: system

The masked data type and size should be compatible with column data type and size. You can check if they are compatible by showing a sample of the masked data.

Table Name: HR.EMPLOYEES Original Data: Not Generated
Column Name: EMPLOYEE_ID Masked Data: Not Generated Show Sample

☐	Big Random Number	Numeric	Random Number (477456-987873849)	SYSMAN
☐	Credit (Master) Card	Character	Credit (Master) Card	SYSMAN
☐	Credit (Master) Card - RN	Character	Credit (Master) Card - RN	SYSMAN
☐	Credit (Visa) Card with Post Checksum	Character	Credit (Visa) Card with Post UDF	SYSMAN
☐	Medium Random Number	Numeric	Random Number (56789-99999999)	SYSMAN
☐	NH Phone Number	Character	NH Phone Number	SYSMAN
☐	NH Phone Number - RN	Character	NH Phone Number - RN	SYSMAN
☐	Random Date	Date	Random Date	SYSMAN
☐	Small Random Number	Numeric	Random Number (1-101)	SYSMAN
☒	Social Security Number	Character	Social Security Number	SYSMAN

Post-Processing Function

Name: Example: sysman.my_checksum_func



User-defined mask formats

Email notification testing

Format Entries

Define masking format by adding one or more format entries of different types.

Type	Description
User Defined Function	Function Name: hr.email_mask

Sample Masked Data

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

- Harry Dean.978448089.Dvrrie@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

Masking Definition: Employee HR mask

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

Name	Employee HR mask	Database	Oemrep_Database	Description
------	------------------	----------	-----------------	-------------

Columns

Owner	Table Name	Column Name	Format
HR	EMPLOYEES	EMPLOYEE_ID	000
HR	EMPLOYEES	FIRST_NAME	000
HR	EMPLOYEES	LAST_NAME	000

Foreign Key Columns

Owner	Table Name	Column Name	Parent Owner	Parent Table	Parent Column
HR	DEPARTMENTS	MANAGER_ID	HR	EMPLOYEES	EMPLOYEE_ID
HR	EMPLOYEES	MANAGER_ID	HR	EMPLOYEES	EMPLOYEE_ID
HR	JOB_HISTORY	EMPLOYEE_ID	HR	EMPLOYEES	EMPLOYEE_ID
OE	CUSTOMERS	ACCOUNT_MGR_ID	HR	EMPLOYEES	EMPLOYEE_ID
OE	ORDERS	SALES_REP_ID	HR	EMPLOYEES	EMPLOYEE_ID

Dependent Columns

Owner	Table Name	Column Name	Parent Owner	Parent Table	Parent Column
HR	MANAGERS	MGR_ID	HR	EMPLOYEES	EMPLOYEE_ID

Referential Integrity Enforcement

Masking Definition: Employee HR mask

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

Name **Employee HR mask** Database **Oemrep_Database** Description

Columns

Owner	Table Name	Column Name	Format
HR	<u>EMPLOYEES</u>	EMPLOYEE_ID	999
HR	<u>EMPLOYEES</u>	FIRST_NAME	999
HR	<u>EMPLOYEES</u>	LAST_NAME	999

Foreign Key Columns

Owner	Table Name	Column Name	Parent Owner	Parent Table	Parent Column
HR	<u>DEPARTMENTS</u>	MANAGER_ID	HR	EMPLOYEES	EMPLOYEE_ID
HR	<u>EMPLOYEES</u>	MANAGER_ID	HR	EMPLOYEES	EMPLOYEE_ID
HR	<u>JOB_HISTORY</u>	EMPLOYEE_ID	HR	EMPLOYEES	EMPLOYEE_ID
OE	<u>CUSTOMERS</u>	ACCOUNT_MGR_ID	HR	EMPLOYEES	EMPLOYEE_ID
OE	<u>ORDERS</u>	SALES_REP_ID	HR	EMPLOYEES	EMPLOYEE_ID

Dependent Columns

Owner	Table Name	Column Name	Parent Owner	Parent Table	Parent Column
HR	<u>MANAGERS</u>	MGR_ID	HR	EMPLOYEES	EMPLOYEE_ID

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 Definition **Impact Report** Schedule Review

Mask: Impact Report

Database: Oemrep_Database Number of Tables: 6
Logged In As: system Columns: 9

Script Generation Summary

Most Serious Message Severity: INFORMATION
Generation Started: Sep 14, 2007 2:12:15 PM
Generation Completed: Sep 14, 2007 2:12:21 PM

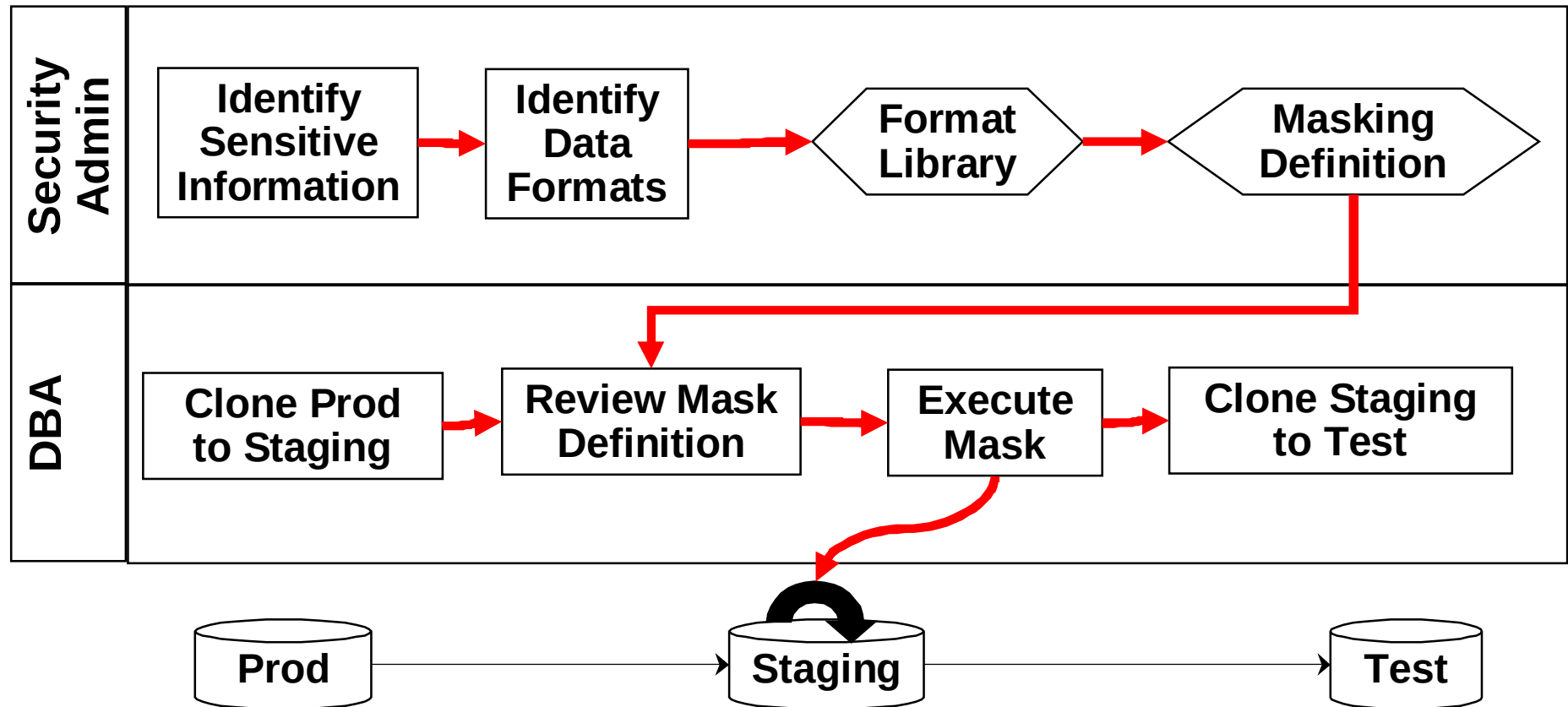
Script Generation Information

The following table provides information about the objects and resources examined during script generation and lists details of any warnings or errors detected.

Object Name	Object Type	Message Severity	Message Type	Message
SYSTEM	TABLESPACE	INFORMATION	Plan	Sufficient free space in Tablespace SYSTEM. Starting Freespace: 312416KB. Ending Freespace: 312632KB. Lowest Freespace: 312456KB.
HR	USER	INFORMATION	Plan	Sufficient tablespace quota for User HR.
OE	USER	INFORMATION	Plan	Sufficient tablespace quota for User OE.

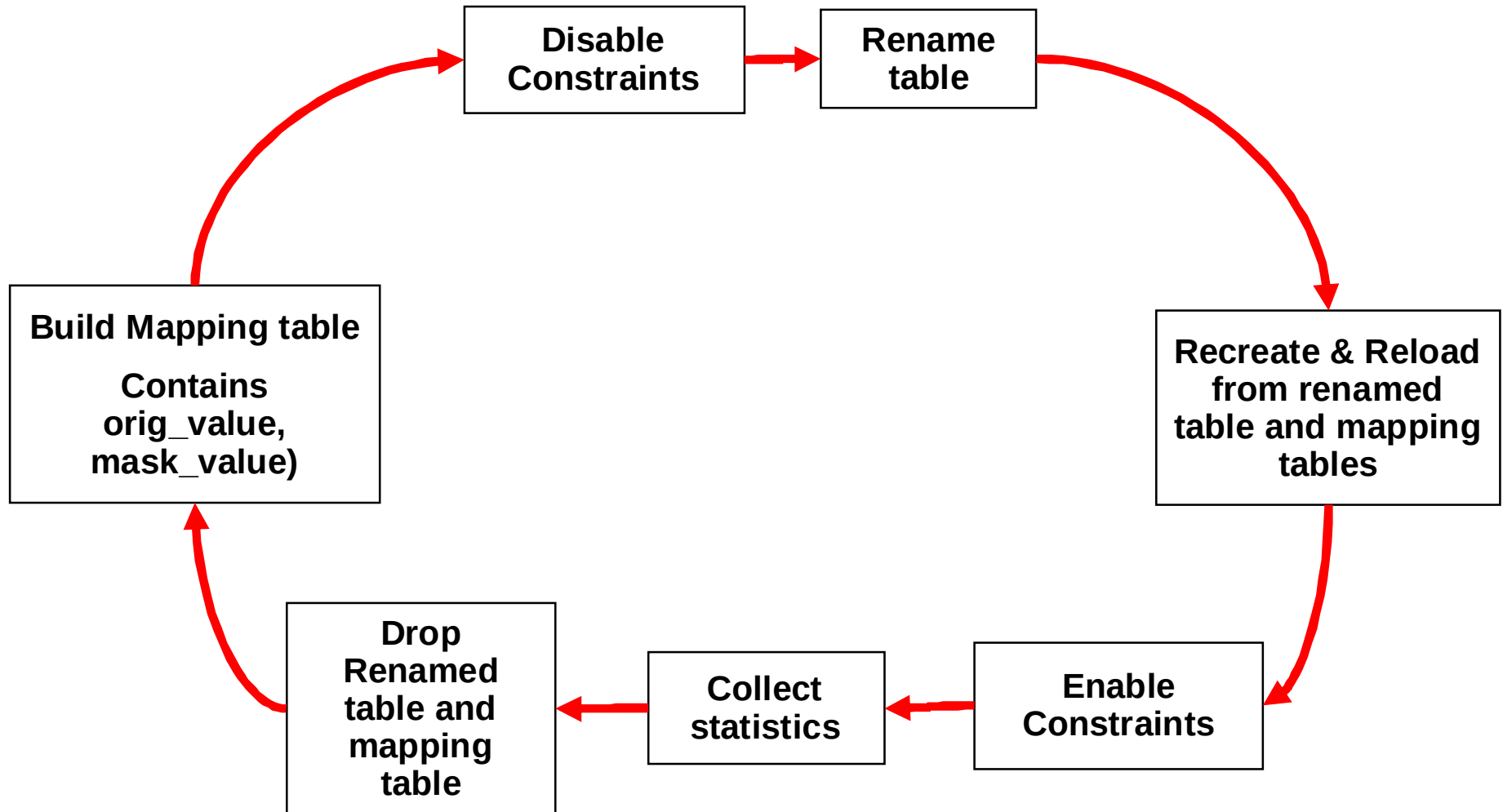


Masking Workflow





Data Masking Internals





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



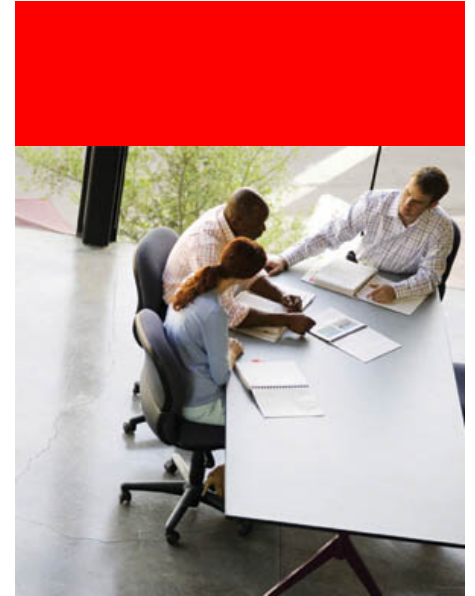
For More Information

<http://search.oracle.com>

Data Masking

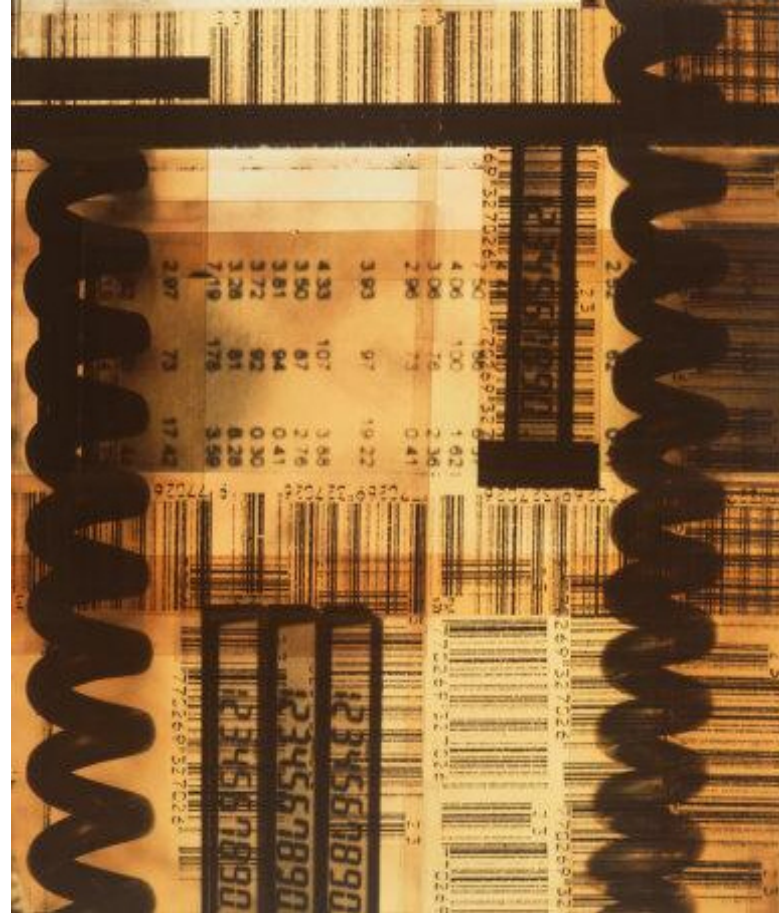


Demonstration



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.



ORACLE®