ORACLE

Oracle Data Masking and Subsetting
Minimize Risk, Maximize Value

Agenda
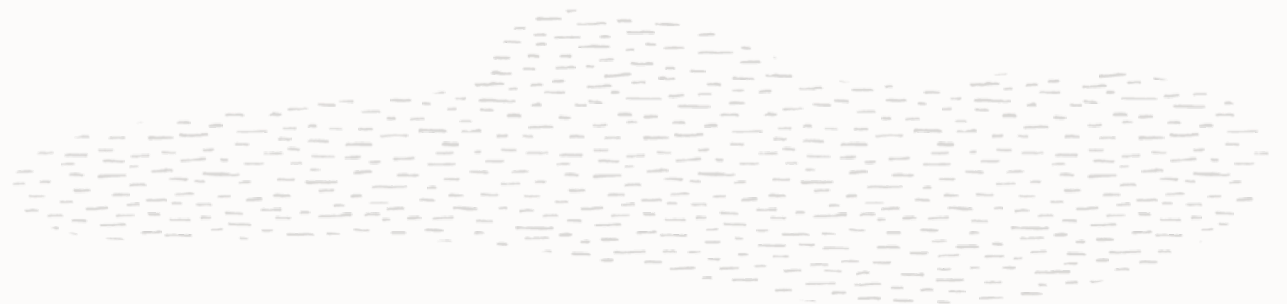
Data is today's capital
"The world's most valuable resource is no longer oil, but data"

—

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

PII Data
Financial Data
Trade Secrets
Competitive Data
Employment Data
Healthcare Data
IT Security Data
Transaction Data
Browsing Data…

Data can be a liability
The scary side of data economy

—

Data breaches are exploding world wide
- 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

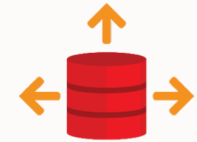Testing   Development   Cloud   Partners   Analytics   Demo   Training   Research   & More...

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
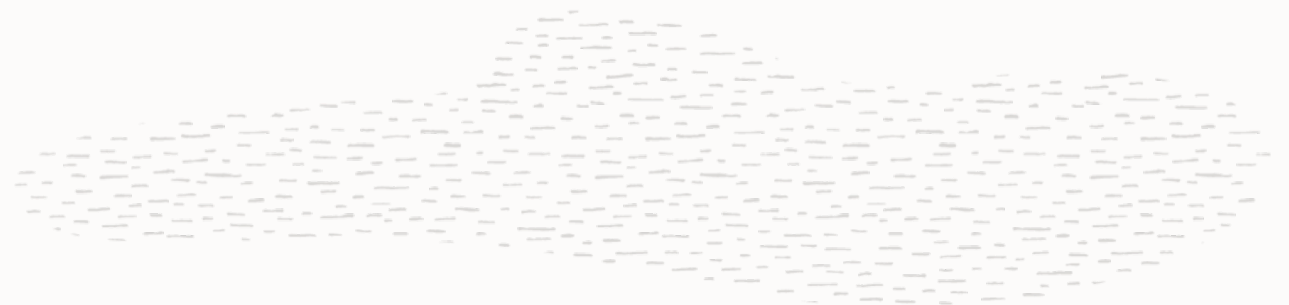- Quickly share data with developers, data scientists, and partners

The Solution?

Your concern

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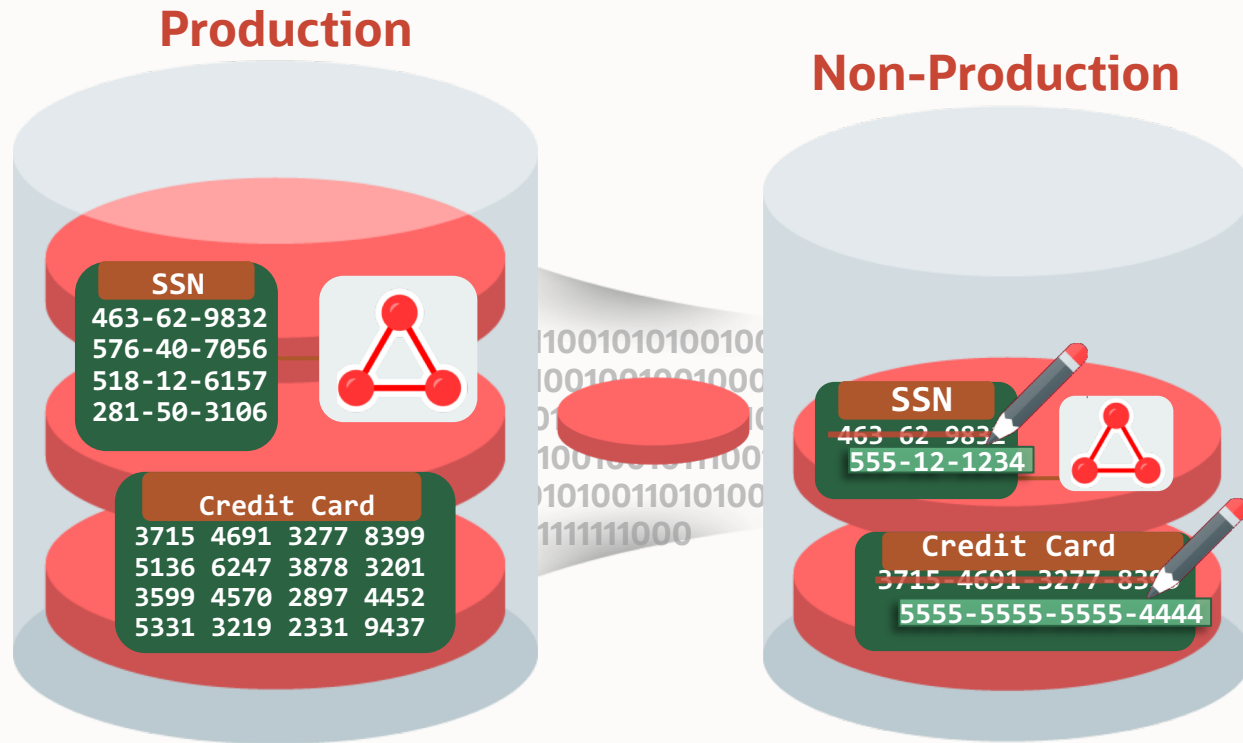- Minimize time and storage costs

Agenda

Oracle Data Masking and Subsetting

Minimize proliferation of sensitive data to non-production environment

ORACLE ENTERPRISE MANAGER **13**^{c}

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
5331 3219 2331 9437

1001010100100
0010001001000
01
100
01
1010011010100
1111111000

Non-Production

SSN
463-62-9832
555-12-1234

Credit Card
3715 4691 3277 839
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
Address
Family Data
Date of Birth
Place of Birth
...

IT

IP Address
User ID
Password
Hostname
GPS location
...

Financial

Credit Card
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
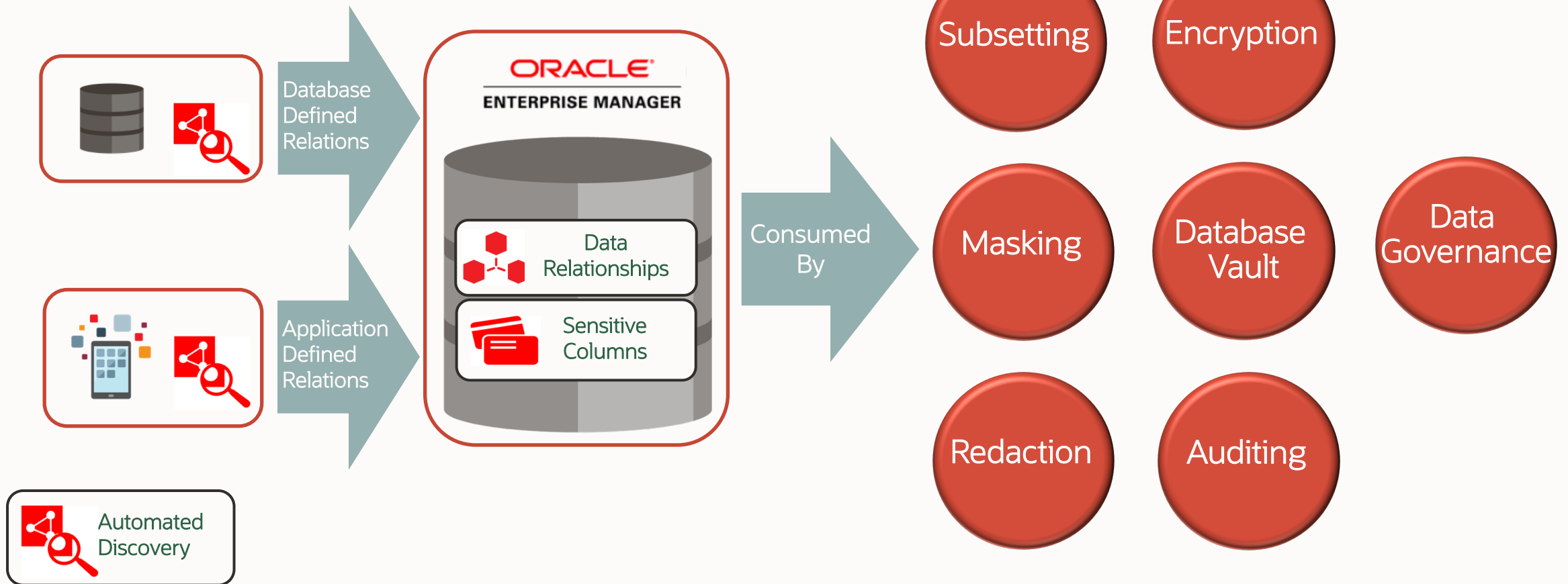Job Title
Department
Hire Date
Income
Stock
...

Academic

College Name
Grade
Student ID
Financial Aid
Admission Date
Graduation Date
Attendance
...

Application Data Modeling
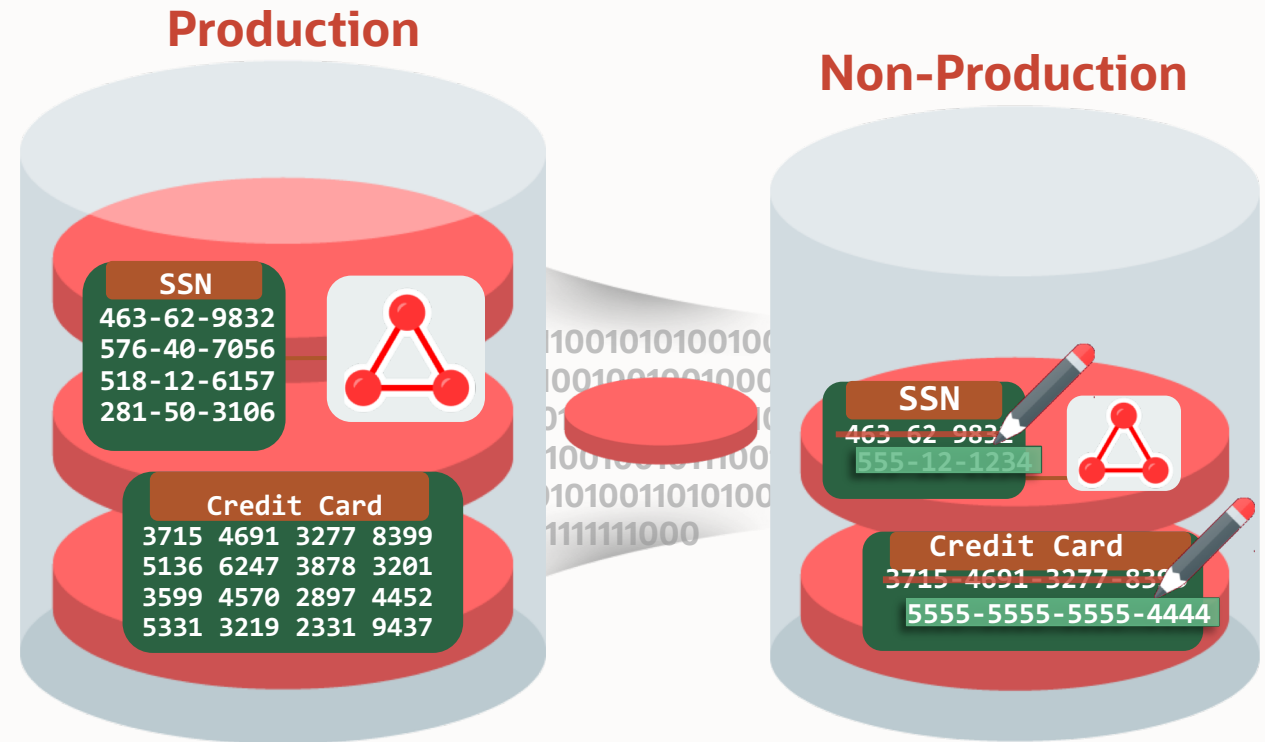## Sensitive Data Discovery

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

Production

Non-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
5331 3219 2331 9437

SSN
463-62-9832
555-12-1234

Credit Card
3715 4691 3277 830
5555-5555-5555-4444

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

Format	Description
American Express Credit Card Number	~10 billion unique American E...
Discover Card Credit Card Number	~10
MasterCard Credit Card Number	~10
Visa Credit Card Number	~10
Generic Credit Card Number	~10
Generic Credit Card Number Formatted	~10
National Insurance Number Formatted	Gen...
Social Insurance Number	~1 b
Social Insurance Number Formatted	~1 b
Social Security Number	~71
Social Security Number Formatted	~71
	~1 b
	~1 b
	~2 b

Array List ▼
- Array List
- Delete
- Encrypt
- Fixed Number
- Fixed String
- Null Value
- Preserve Original Data
- Random Dates
- Random Decimal Numbers
- Random Digits
- Random Numbers
- Random Strings
- Shuffle
- SQL Expression
- Substitute
- Substring
- Table Column
- Truncate
- User Defined Function

Sample Masked Data
Samples are generated using defined format
- 3472105015722069   Refresh
- 3749455677707248
- 3490749344336998
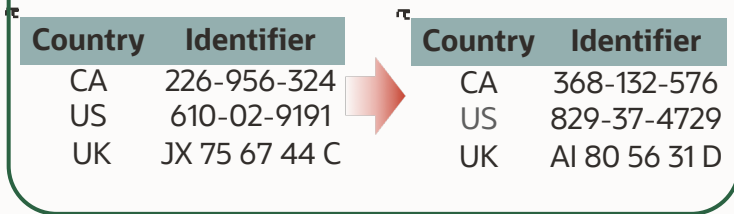- 3782460947413526
- 3452029369341892

Data Masking
Masking transformations to meet diverse business use cases

Conditional masking
Masks rows differently based on condition
Example: Mask national identifiers based on country

Deterministic masking
Masks data to the same consistent values across multiple databases or masking jobs
Example: Mask employee identifiers consistently across schemas and databases

Compound masking
Ensures masked values across related columns retain the same relationship
Example: Mask address fields such as state, postal code, and country as a group

Format preserving
Masks data while preserving its format such as length and special characters
Example: Mask tax identifiers while preserving spaces and hyphens

Reversible masking
Encrypts and decrypts data using cryptographic key
Example: Unmask data after receiving the processed data from a partner

Shuffling
Shuffles the values within a column
Example: Shuffle age of employees in a organization

Perturbation
Generates random values within a user-provided range
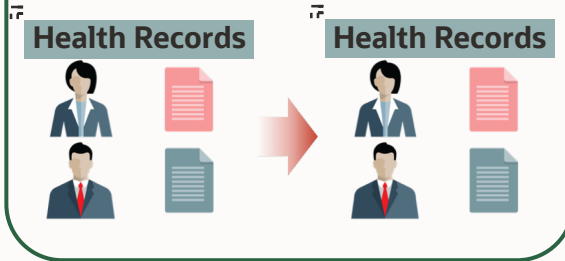Example: Generate random dates within a specified data range
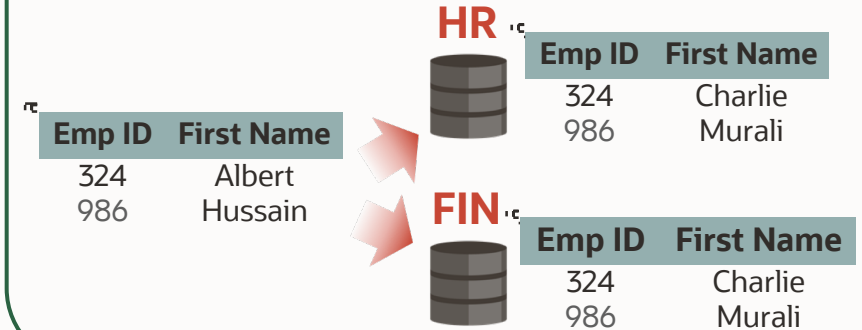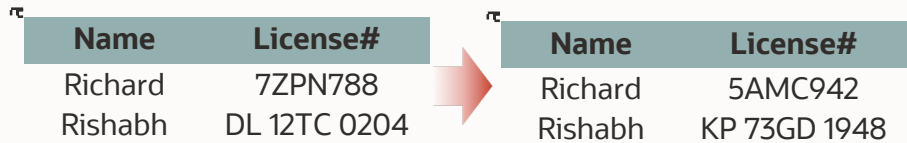
Data Masking
Examples

Mask based on conditions

Country	Identifier
CA	226-956-324
US	610-02-9191
UK	JX 75 67 44 C

Country	Identifier
CA	368-132-576
US	829-37-4729
UK	AI 80 56 31 D

Shuffle records

Health Records → Health Records

Generate deterministic output

Emp ID	First Name
324	Albert
986	Hussain

HR

Emp ID	First Name
324	Charlie
986	Murali

FIN

Emp ID	First Name
324	Charlie
986	Murali

Generate random values while preserving format

Name	License#
Richard	7ZPN788
Rishabh	DL 12TC 0204

Name	License#
Richard	5AMC942
Rishabh	KP 73GD 1948

Mask operating system files stored as LOBs

LOB
3178973456
6509876745

```
Search : [0-9]{10}
Replace : *
```

LOB
****
****
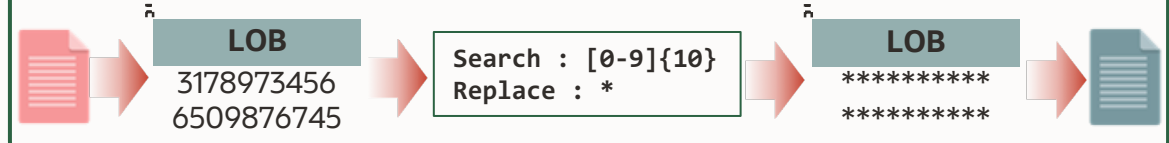
Data Masking
Masking templates

—

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
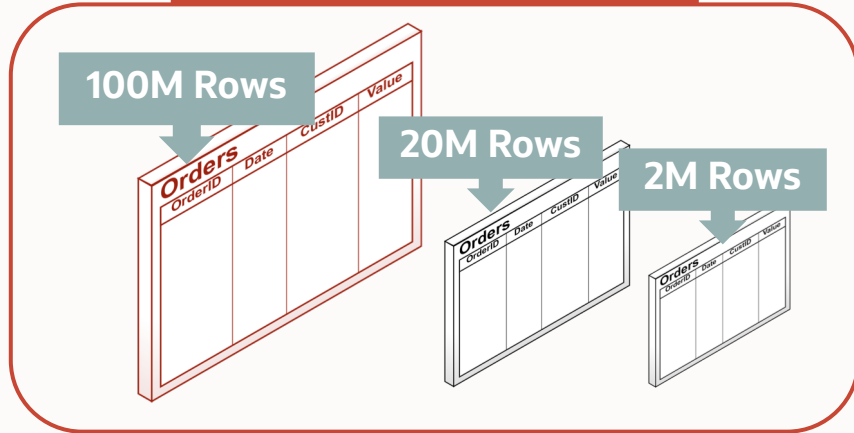
Perform research and analysis on a subset of data

Extract subset of data as part of e-discovery requests

Data Subsetting
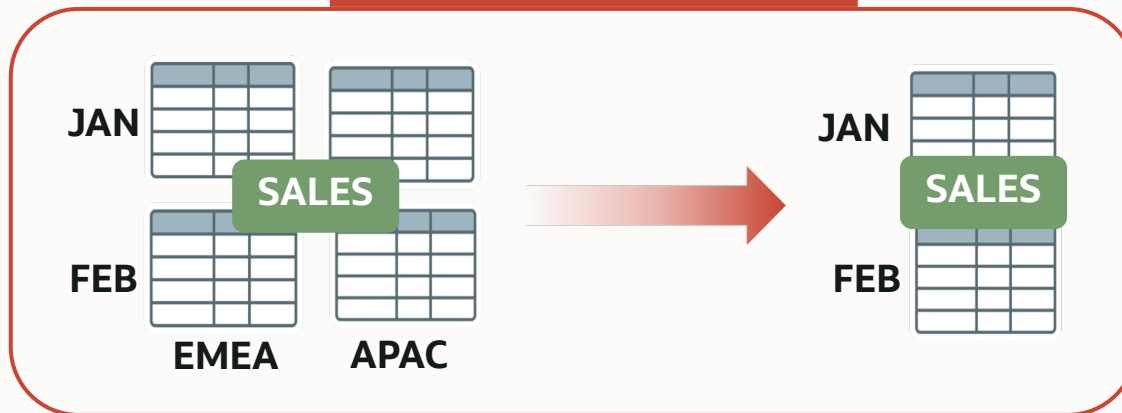Goal or condition based subsetting

Relative Table Size

100M Rows

20M Rows

2M Rows

Orders | OrderID | Date | CustID | Value

Condition Based

SALES

Extract ASIA Sales

ASIA

Table Partitions

JAN

FEB

SALES

EMEA   APAC

JAN

FEB

SALES

Data Subsetting
Preview and validate results

| Applications | Table Rules | Rule Parameters | **Space Estimates** | Pre/Post Subset Script |

Impact of subset rules on tables are displayed below. The values shown here are based on estimates and may not be accurate.

View ▾   Refresh...

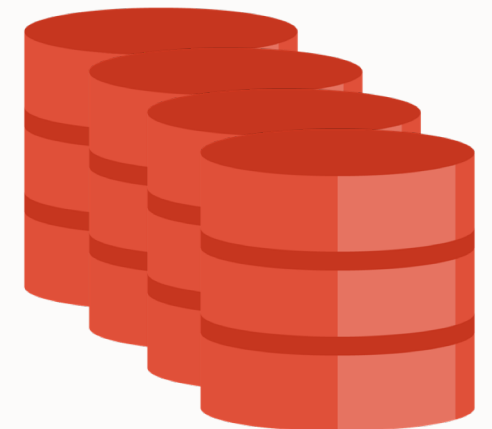| Name	Table Rule	Source Size		Estimated Subset Size		
		MB	Rows	MB	Rows	%
▽ Applications and Tables		912.1908	8661245	483.5967	4438883	53.01
▽ TDM(TDM)		912.1908	8661245	483.5967	4438883	53.01
H_LINEITEM		606.6597	6001215	246.1027	2434504	40.57
H_ORDER	o_custkey in (select c_custkey from tdm.h_custo...	148.7732	1500000	88.6328	893637	59.58
H_PARTSUPP		109.1003	800000	109.1003	800000	100
H_CUSTOMER		24.0326	150000	16.1358	100712	67.14
H_PART		22.316	200000	22.316	200000	100
H_SUPPLIER		1.3065	10000	1.3065	10000	100
H_NATION		0.0021	25	0.0021	25	100
H_REGION		0.0004	5	0.0004	5	100

Agenda
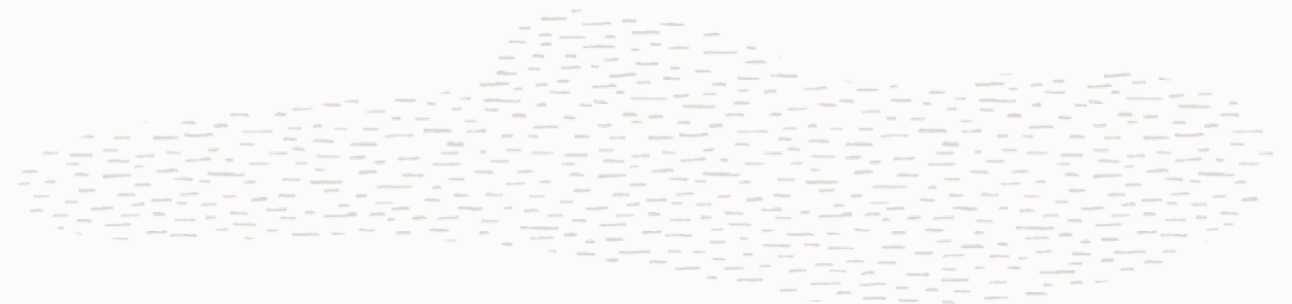
Streamlined Management
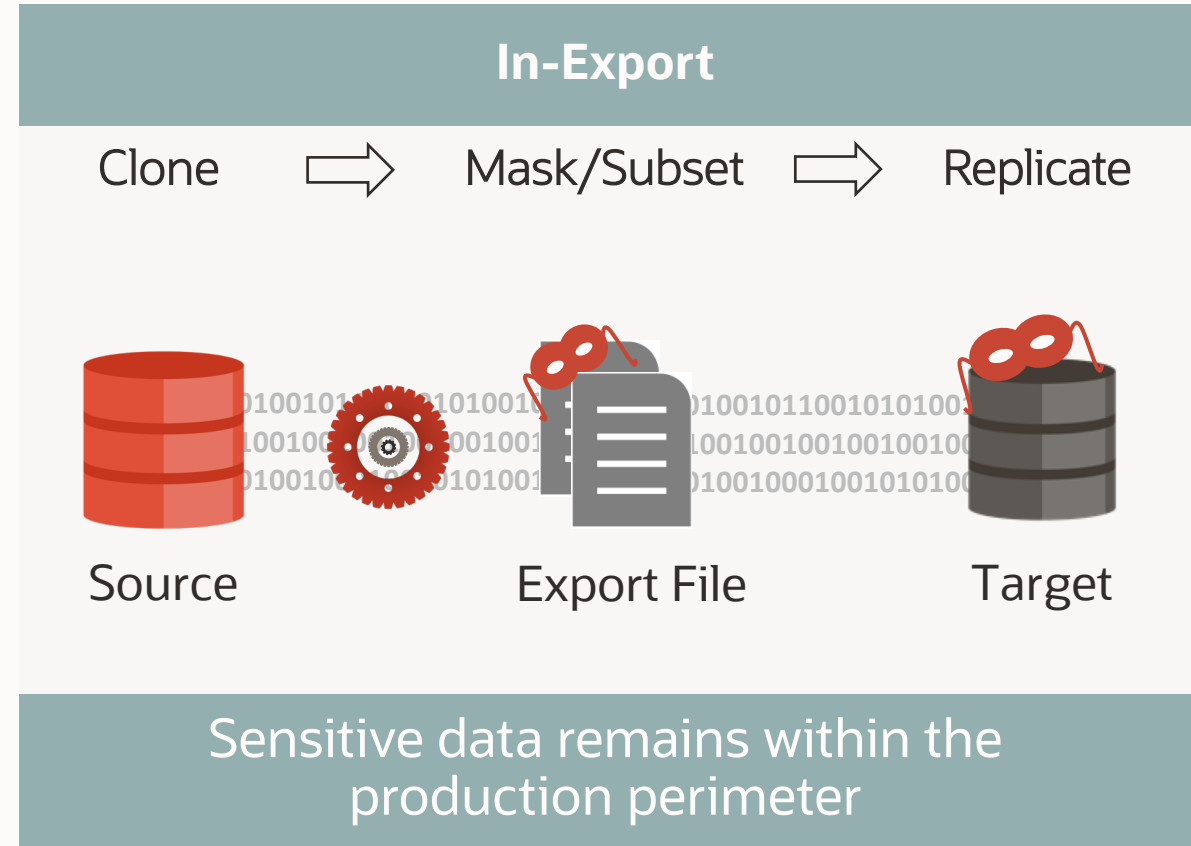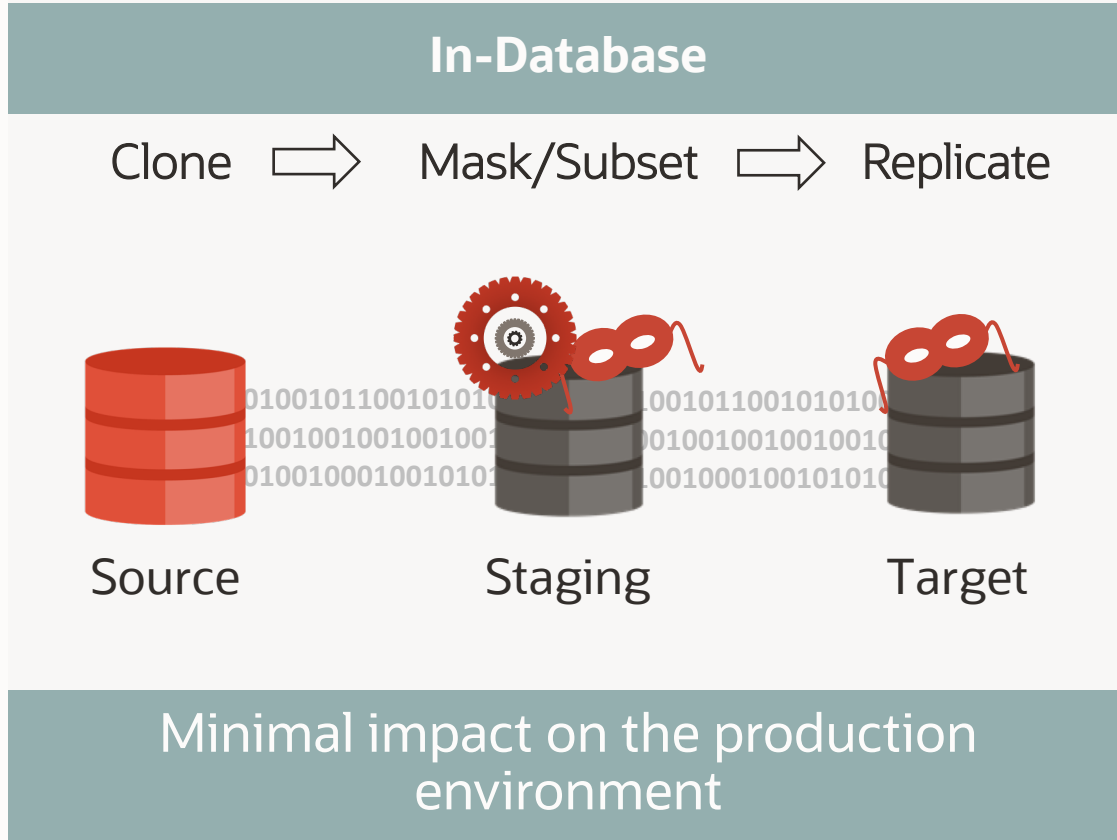
Preinstalled with Oracle Enterprise Manager

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

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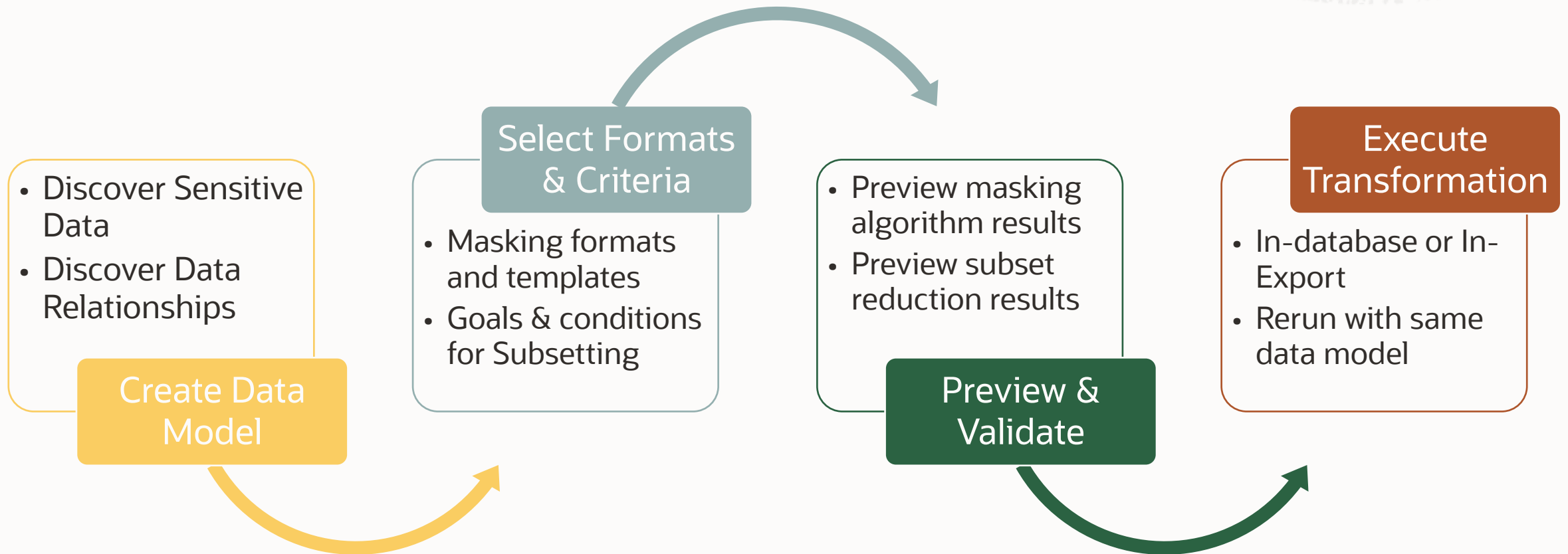
Flexible Deployment Options

In-Database
Clone ⇨ Mask/Subset ⇨ Replicate
Source   Staging   Target
Minimal impact on the production environment

In-Export
Clone ⇨ Mask/Subset ⇨ Replicate
Source   Export File   Target
Sensitive data remains within the production perimeter

Data Masking Integrations

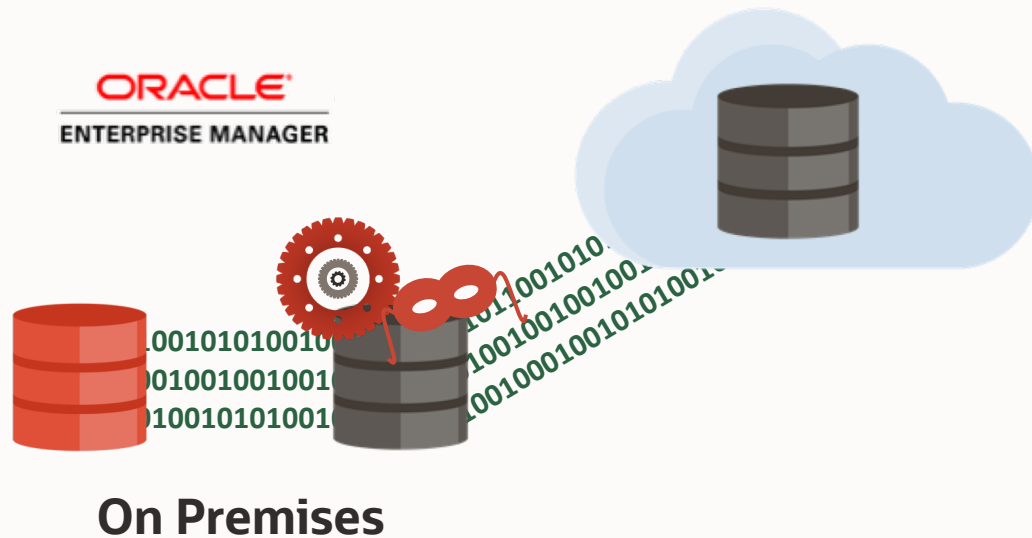| Integration Point	Benefit	
Database Lifecycle Management Pack	Integrated clone and mask	✓
Cloud Management Pack	Integrated snap-clone and mask	✓
Real Application Testing Pack	Mask in database workload captures	✓
Oracle Data Integrator	Mask incrementally captured data	✓

Implementation Methodology

Create Data Model
- Discover Sensitive Data
- Discover Data Relationships

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

Mask on Premises and Upload to the Cloud
Oracle Database Cloud Service (PaaS)

Clone ➡ **Mask/Subset** ➡ **Upload**

Extract ➡ **Mask/Subset** ➡ **Upload**

ORACLE®
ENTERPRISE MANAGER

On Premises

ORACLE®
ENTERPRISE MANAGER

On Premises

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

Clone & Mask PDB to Cloud

On Premises
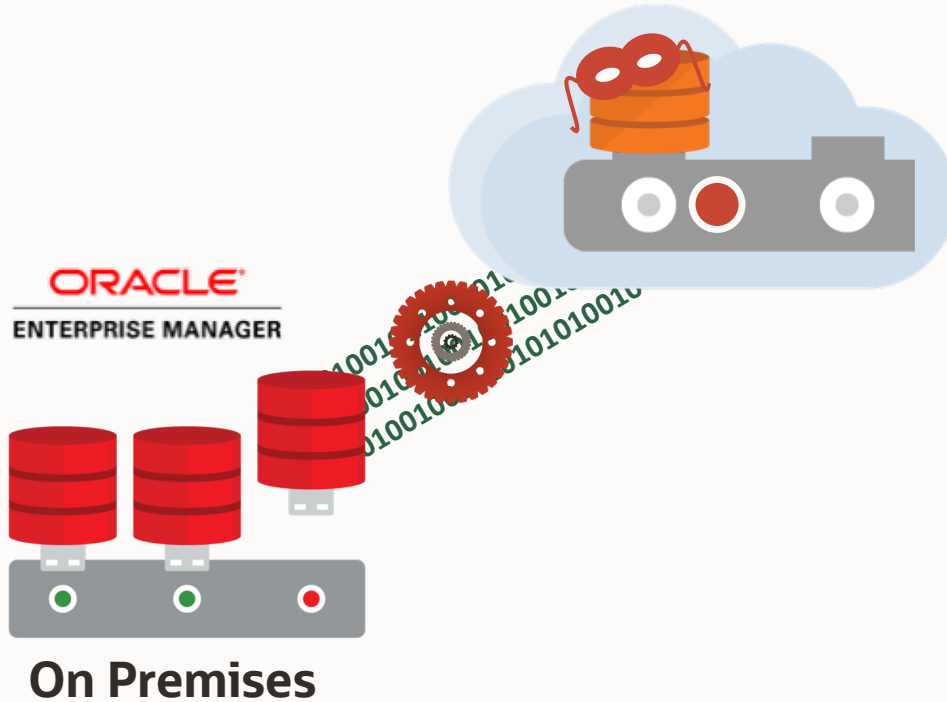
Mask/Subset in the Cloud

On Premises
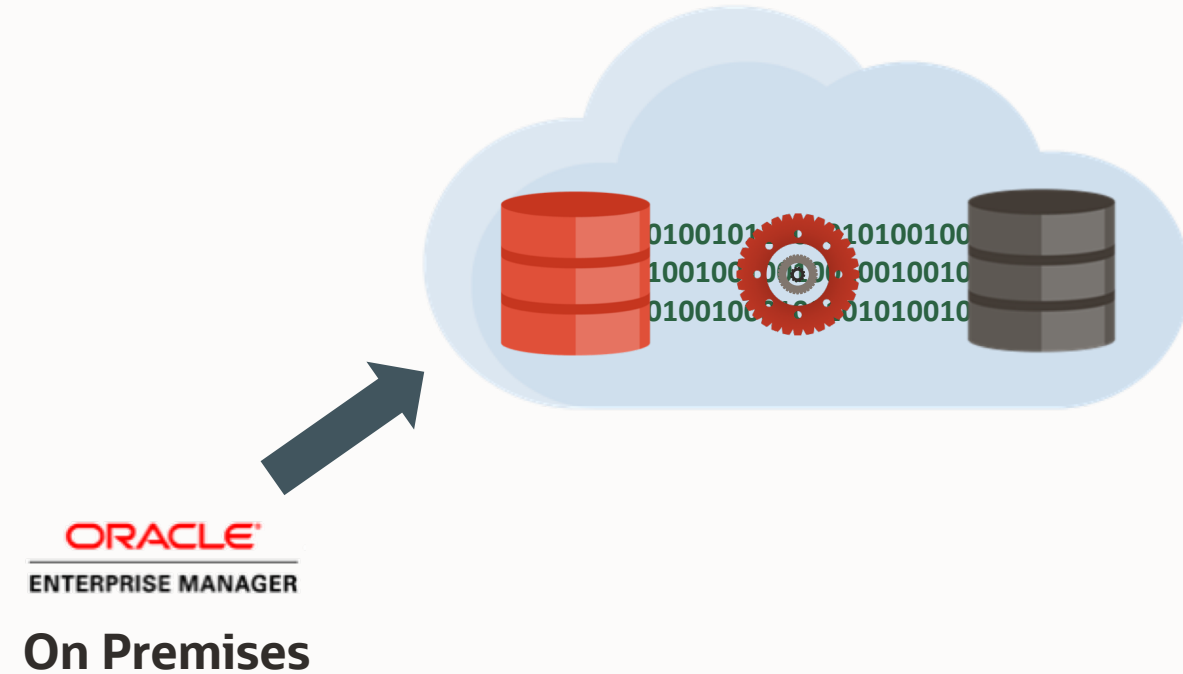
Masking in Oracle Fusion Cloud Service

1 Customer requests to mask sensitive data in dev/test

HCM Cloud **+** Cloud ERP / CX Cloud Suite

Production

Dev/Test

0100101100101010
1001001001001001
0100100010010101
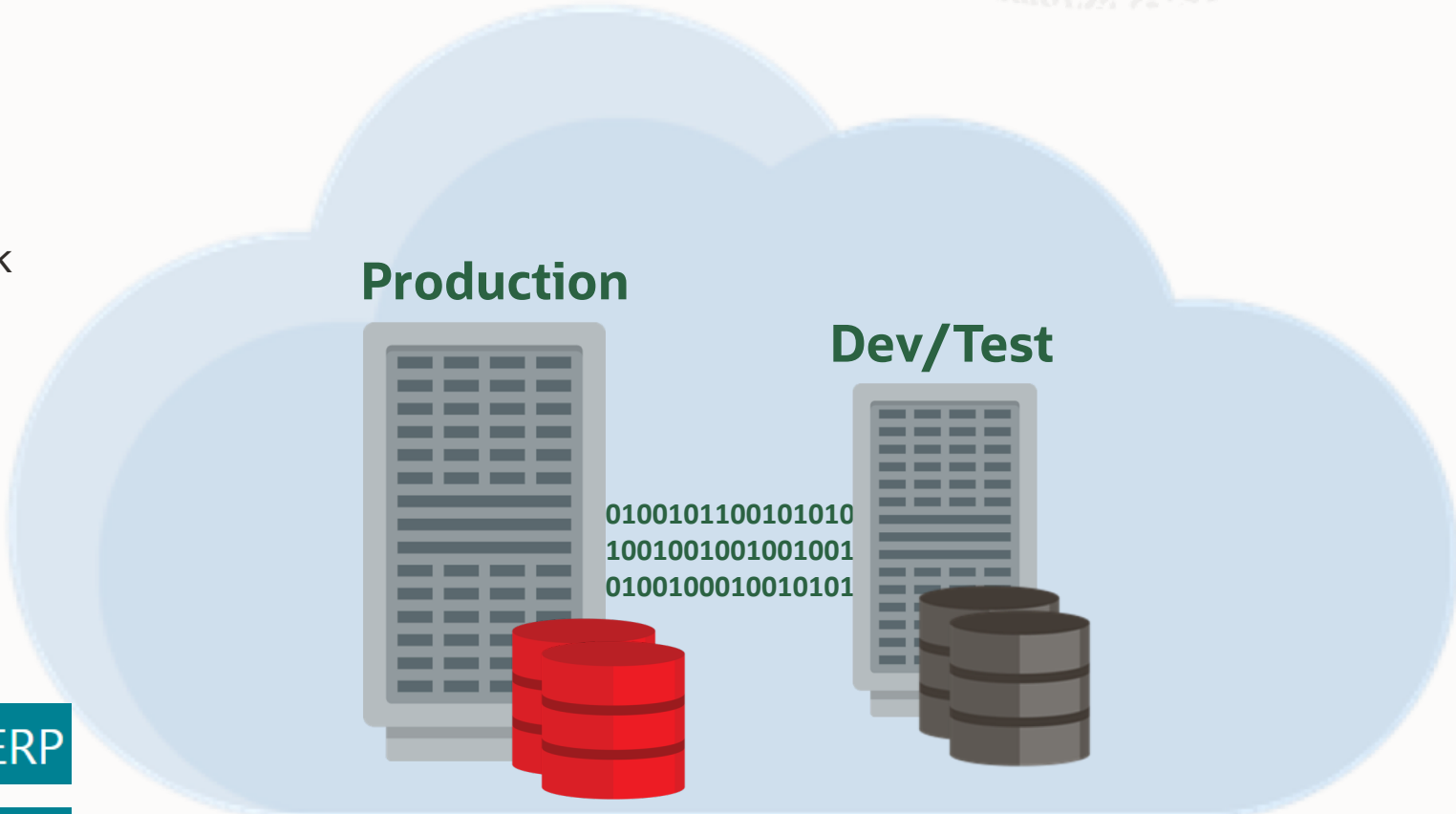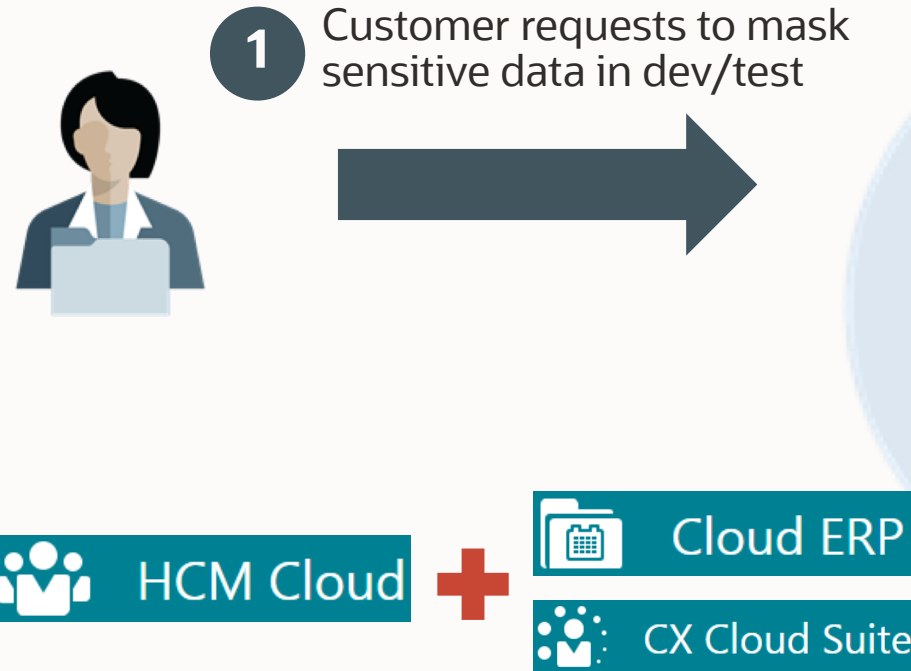
2 Oracle Managed Cloud Services team masks the sensitive data using automated procedures

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.