ORACLE

Oracle Data Masking and Subsetting

Frequently Asked Questions

July, 2024 Version 13c
Copyright © 2024, Oracle and/or its affiliates

Public

ORACLE

Table of contents

ORACLE

Introduction

Growing security threats and ever-expanding privacy regulations have made it necessary to limit the exposure of sensitive data. Copying production data for non-production purposes such as development and data analytics proliferates sensitive data, expanding the security and compliance boundary and increasing the likelihood of data breaches. Oracle Data Masking and Subsetting provides a flexible solution that discovers, masks, and subsets sensitive data, allowing organizations to safely share data across their non-production environments.

Product Overview

What is Data Masking and Subsetting?

Data masking (sometimes called static data masking) replaces sensitive data such as credit card numbers with fictitious yet realistic-looking data. Data subsetting is the process of retaining or extracting a selected portion of a data set from a larger database.

Why do I need to mask and subset data?

Copying production data to non-production, outsourced, partner, and cloud environments for test, development, and other purposes proliferates sensitive information such as credit card numbers and social security numbers, increasing the risk of a data breach as non-production environments are generally not as protected or monitored as diligently as production environments. For this reason, data privacy standards such as PCI-DSS recommend rendering sensitive production data unreadable when used for test and development.

Subsetting extracts only the necessary information from a large database for sharing with internal and external teams, reducing the resources required to store and manage that data in test and development. Masking and subsetting sensitive data in non-production environments helps improve security and minimize compliance and infrastructure costs.

How does masking improve security and minimize compliance costs?

Masking sensitive data in test and development environments reduces the overall compliance boundary, restricting it to only production environments. Rendering data unreadable in these environments limits the risk of a data breach and helps minimize compliance costs.

How does subsetting minimize infrastructure costs?

Reducing the size of non-production databases by subsetting data minimizes the storage cost. Using smaller portions of data for test and development environments also reduces the time and expense of transferring the data and computing costs.

Why are data masking and subsetting important for cloud computing?

Organizations understand the advantage of leveraging a cloud platform for test and development. However, they are concerned about uploading sensitive on-premises production data to the cloud because of data privacy and

ORACLE

compliance reasons. Other concerns are the storage cost associated with the cloud platform and the network cost due to data transfers.

Oracle Data Masking and Subsetting addresses these concerns by enabling cloud users to mask sensitive data on-premises before uploading it to the cloud. The product helps reduce storage and network costs by extracting a subset of production data for upload to the cloud.

Components and Features

What are the main components of Oracle Data Masking and Subsetting?

The main components are:

- **Application Data Modeling** provides automated procedures to discover sensitive columns and parent-child relationships. The discovery results are stored as an application data model, which is reusable across multiple databases.

- **Masking Format Library** provides a comprehensive set of predefined masking formats to mask sensitive data such as credit card numbers, national identifiers, and phone numbers. It also allows creating new masking formats to meet domain-specific requirements.

- **Data Masking** assists in mapping masking formats to discovered sensitive columns, creating reusable masking scripts. It also provides a workflow to mask data.

- **Data Subsetting** helps create reusable goal/condition-based subsetting rules on a database. It also provides a workflow to generate subsets.

How does the product preserve the relational integrity of the data in an application?

Oracle Data Masking and Subsetting does the following to minimize the disruption of applications post-masking and subsetting:

- The product uses automated discovery procedures to gather referential integrity or parent-child relationships between the columns before the masking and subsetting process.

- During the masking and subsetting process, parent and child columns are processed consistently to preserve the integrity between these columns.

Can masking support multi-byte or international characters?

Several masking options support multi-byte or international characters, such as UTF-8. The suitable masking formats include Array List, Shuffle, Substitute, Table Column, and User Defined Function.

Does the product include predefined masking formats?

Yes, Oracle Data Masking and Subsetting provides out-of-the-box masking formats covering a broad range of sensitive data, such as national identifiers of multiple countries, credit card numbers of various vendors, phone numbers, and more.

ORACLE

Which masking techniques are supported by the product?

Some options include generating fixed/random characters or numbers, replacing them with null values, substituting data from a random list or table column, and SQL or regular expression-based masking. You also have several advanced options to meet complex business requirements, such as:

- **Shuffle Masking** randomly shuffles data within a column. For example, columns containing salaries can be shuffled to break the employee-salary mapping.

- **Encryption** encrypts sensitive data using a cryptographic key while preserving the data's format. It's a reversible masking option, and you can decrypt your data using the same key. This feature is useful when masked data sent to a third party has to be merged with further updates.

- **Format Preserving Randomization** randomizes the data while preserving the input length, character position, case, and special characters.

- **Conditional Masking** masks column data using different masking formats based on user-defined conditions. For example, in a column, the US identifiers can be masked using the Social Security Number format and the UK identifiers using the National Insurance Number format.

- **Compound Masking** masks related columns as a group, ensuring the masked data across the related columns retain the same relationship. For example, address fields such as city, state, and postal codes can be masked consistently.

- **Deterministic Masking** generates consistent masked output for a given input across application schemas and databases.

- **User-defined PL/SQL Masking** enables you to define custom masking logic or migrate your existing masking scripts.

What assurances does the product provide for the integrity of the masked and subsetted data?

Each masking format has built-in logic to validate the generated masked output. For example, all credit card number masking formats perform the Luhn check on the masked credit card numbers. Also, the product provides a preview option to validate the defined masking and subsetting criteria before running masking and subsetting scripts.

Can I migrate in-house masking scripts to Oracle Data Masking and Subsetting?

Yes, the product supports the migration of existing masking scripts for generating custom data types with user-defined PL/SQL masking functions.

Which subsetting techniques are supported by the product?

Oracle Data Masking and Subsetting simplifies the task of subsetting through its goal or condition-based subsetting techniques. A goal can be a relative table size, such as extracting a 1% subset of a table containing 10 billion rows. A condition can be based on time, such as discarding all user records created before a particular year.

ORACLE

A condition can also be based on region, such as extracting only Asia Pacific information to support new application development.

Does Oracle Data Masking and Subsetting work with packaged applications like Oracle E-Business Suite and Oracle Fusion Applications?

As Oracle Data Masking and Subsetting is a database-centric solution, it works for all supported databases regardless of the application. However, care is required when setting up data models and masking and subsetting definitions to avoid misconfigurations that could break complex applications. Oracle E-Business Suite and Oracle Fusion Applications provide prepackaged application data models and masking definitions for use with Oracle Data Masking and Subsetting.

Can I mask and subset databases running in Oracle Cloud?

Yes, you can mask and subset databases in the Oracle Cloud. Oracle Data Masking and Subsetting for cloud databases works much like on-premises databases. Oracle Data Safe, an OCI-native cloud service that supports sensitive data discovery and masking, is also available for Oracle Cloud users.

Can I mask and subset data in non-Oracle relational databases?

Oracle Data Masking and Subsetting can mask and subset data in DB2, Informix, SQL Server, Sybase, MySQL, and Teradata. These sources are supported through Oracle Transparent Gateways, which are included with the product.

DEPLOYMENT AND ADMINISTRATION

How do I download and install Oracle Data Masking and Subsetting?

Oracle Data Masking and Subsetting is pre-installed with Oracle Enterprise Manager. To use Oracle Data Masking and Subsetting, you must have a valid license for the pack.

What are the different ways to mask and subset data?

The product provides two modes to mask and subset data:

- **In-Database Masking and Subsetting:** The target production data is first copied (cloned) to a separate location. Oracle Data Masking and Subsetting operates on the cloned data. After processing is complete, the resulting masked data can be cloned and distributed for non-production use.

- **In-Export Masking and Subsetting:** The masking and subsetting rules are applied while the data is extracted from the target database, and the resulting data is written to Oracle Data Pump dump files. In this mode of operation, Oracle Data Masking and Subsetting can run directly on the production system and unmasked data does not leave production. After processing is complete, the dump file containing masked data can be imported into non-production databases.

How does Oracle Data Masking and Subsetting optimize performance?

Oracle Data Masking and Subsetting implements several database optimizations to deliver high-speed masking and subsetting during in-database operations. It leverages Oracle Data Pump for in-export operations to achieve high performance during masking and subsetting.

ORACLE

MORE INFORMATION

How does Oracle Data Masking and Subsetting compare to other similar Oracle offerings?

ORACLE SOLUTIONS	DESCRIPTION
Oracle Data Masking and Subsetting	Software for creating masked and subsetted copies of production data for use in non-production environments such as testing and development databases.
Oracle Data Safe (Data Discovery/ Data Masking)	Discover and mask sensitive data with a cloud service that supports Oracle Databases everywhere: in the Oracle Cloud, on-premises, and third-party clouds.
Oracle Data Redaction	Redacts sensitive data from query results before display through client applications. Enforces redaction at runtime, with low overhead, and according to conditions set in policies.
Oracle Label Security	Implements Multi-Level Security (MLS), enabling rows with differing sensitivity to reside in the same table. Explicitly labels rows with group, compartment, and sensitivity levels then matches them with user labels.

Where can I find more information on Oracle Data Masking and Subsetting?

For more information, such as product data sheets, tutorials, documentation, customer references, and blogs, please visit the Oracle Data Masking and Subsetting page on Oracle.com.

https://www.oracle.com/security/database-security/data-masking/

Explore the Data Masking and Subsetting workshop on Oracle LiveLabs to try the product at no cost.

https://apexapps.oracle.com/pls/apex/r/dbpm/livelabs/view-workshop?wid=704

ORACLE

Connect with us

Call +**1.800.ORACLE1** or visit **oracle.com**. Outside North America, find your local office at: **oracle.com/contact**.

blogs.oracle.com   facebook.com/oracle   twitter.com/oracle