Oracle Data Masking and Subsetting

Frequently Asked Questions

Public

ORACLE
Popular Topics

ORACLE

Abstract

The growing security threats and ever-expanding privacy regulations have made it necessary to limit 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 the data to be safely shared across non-production environments. Data Masking or Static Data Masking is the process of replacing sensitive data such as credit card numbers with fictitious yet realistic-looking data. Data Subsetting (or Test Data Management) is the process of retaining or extracting a smaller portion of a data set from a large database.

Product Overview

What is Data Masking and Subsetting?

Data Masking or Static Data Masking is the process of replacing sensitive data such as credit card numbers with fictitious yet realistic-looking data. Data Subsetting (or Test Data Management) is the process of retaining or extracting a smaller portion of a data set from a large 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 spreads 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 production environments. Due to this reason, data privacy standards such as PCI-DSS do not recommend using sensitive production data for test and development.

Subsetting extracts only the necessary information from a large database for sharing with internal and external teams, reducing the risk of accidental sensitive data disclosure to unauthorized users.

Masking and subsetting sensitive data in non-production environments helps improve security and minimize compliance and infrastructure costs.

How does masking and subsetting improve security and minimize the compliance cost?

Masking and subsetting sensitive data in test and development environments reduces the overall compliance boundary and the risk of a data breach, thereby, minimizing the compliance costs.

How does the product minimize infrastructure cost?

Reducing the size of non-production databases by subsetting data minimizes the storage cost. Using smaller portions of data for test and development environments minimizes the data provisioning cost.

ORACLE

How does Oracle Data Masking and Subsetting compare to Data Redaction?

Oracle Data Masking and Subsetting (DMS) is used for static masking, where sensitive data is permanently replaced in non-production environments. It supports data discovery, masking, and subsetting for testing, development and, analytics. Data Redaction is for dynamic masking in production environments, where data is masked at runtime based on user roles, without modifying the stored data. Use DMS for persistent, shareable masked data and Redaction for real-time, role-based data protection.

What is the significance of data masking and subsetting in the context of a cloud?

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 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 masking sensitive data on premises before uploading it to the Oracle Cloud. The product helps reduce the storage and network costs by extracting a subset of production data that can be uploaded 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 provides the capability to create 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 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 prior to 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.

How does masking work with 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 multiple vendors, phone numbers, and more.

Which masking techniques are supported by the product?

There are options such as generating fixed/random characters or numbers, replacing with null value, substituting data from a random list or a 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 the sensitive data using a cryptographic key while preserving the format of the data. It's a reversible masking option as you can decrypt your data using the same key. It's useful when masked data sent to a third party has to be merged back along with further updates.

- **Format Preserving Randomization** randomizes the data while preserving the input length, position and case of characters, 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 prior to the execution of masking and subsetting scripts.

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

Yes, the product supports migration of existing masking scripts for custom data types through 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 1% subset of a table containing 10 billion rows. A condition can be based on time, for example, discarding all user records created prior to a particular year. A condition can also be based on region, for example, extracting Asia Pacific information for a new application development.

Does Oracle Data Masking and Subsetting works with applications such as Oracle E-Business Suite, Oracle Fusion Applications, Oracle PeopleSoft, and SAP?

As Oracle Data Masking and Subsetting is a database-centric solution, it works for all supported databases regardless of the application. However, care must be taken when setting up data models and masking and subsetting definitions to avoid misconfigurations that could break complex applications. Certain applications provide pre-packaged data models and masking definitions to avoid the possibility of application breakage and to reduce the customer effort required. Oracle E-Business Suite and Oracle Fusion Applications provide prepackaged 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 Oracle Database Cloud Service (DBCS). Oracle Data Masking and Subsetting for cloud databases works much like it does for on-premises databases. Oracle Data Masking and Subsetting license is included in DBCS High Performance, Extreme Performance, and Exadata Service.

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

Yes, Oracle Data Masking and Subsetting can mask and subset data in DB2, Informix, SQL Server, Sybase, MySQL, and Teradata.

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.

ORACLE

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 being extracted from the target database, and the resulting data is written to Oracle Data Pump dump files. Oracle Data Masking and Subsetting operates directly on the production system – unmasked data does not leave production. After processing is complete, the sanitized dump file can be imported into non-production databases.

What is the performance of the product?

Oracle Data Masking and Subsetting delivers high performance by leveraging the Oracle Database Kernel and Data Pump. It uses CTAS**,** parallel execution**,** and temporarily disables constraints and triggers to reduce overhead. These are restored after processing, and optimizer statistics are recomputed to ensure continued performance, enabling efficient handling of large-scale data.

MORE INFORMATION

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

ORACLE SOLUTIONS	DESCRIPTION
Oracle Data Redaction	Dynamically masks sensitive data at runtime, without altering the stored values. Ideal for production environments where data needs to be protected but still accessed
Oracle Data Masking and Subsetting	Integrated with Enterprise Manager, On-prem/hybrid solution for large-scale, high-performance discovery, static masking and subsetting. Best for deep customization and complex enterprise use cases
Oracle Data Safe (Data Discovery/ Data Masking)	Cloud-native, agentless solution for discovery and static masking. Ideal for quick setup, centralized control, and cloud databases

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

For more information such as product data sheet, tutorials, documentation, customer references, and blog, please visit the following Oracle Data Masking and Subsetting page on Oracle Technology Network.

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

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