

# Oracle Data Masking and Subsetting

## Frequently Asked Questions (FAQ)

SEPTEMBER 2016





## Product Overview

### **Q: What is Data Masking and Subsetting?**

**A:** Data Masking or Static Data Masking is the process of replacing a sensitive value such as credit card number with a fake yet realistic looking credit card number. Data Subsetting or Test Data Management is the process of retaining or extracting a smaller portion of a data set from a large database.

### **Q: Why do I need to mask and subset data?**

**A:** Copying production data to non-production, outsourced, partner, and cloud environments for test and development purposes spreads sensitive information such as credit card numbers and social security numbers increasing the risk of an attack or a data breach as generally non-production environments are 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 test and development environments helps to improve security minimize compliance and infrastructure costs.

### **Q: How does masking and subsetting improve security and minimize the compliance cost?**

**A:** 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.

### **Q: How does the product minimize infrastructure cost?**

**A:** Reducing the size of test/development databases by subsetting data minimizes storages cost. Using smaller portions of data for test and development environments minimizes data provisioning cost.

### **Q: What is the significance of data masking and subsetting in the context of a Cloud?**

**A:** Organizations understand the advantage of leveraging a cloud platform for test and development. However, they are concerned about uploading sensitive on-premise production data to the cloud for data privacy and compliance reasons. Other concerns are storage cost associated with the cloud platform and network cost due to the data transfers.

Oracle Data Masking and Subsetting addresses these concerns by masking sensitive data on-premise before uploading the data to the Oracle Cloud. The product helps to reduce the storage cost and network cost by extracting a subset of production data that can be uploaded to the cloud.



## Components and Features

### **Q: What are the main components of Oracle Data Masking and Subsetting?**

**A:** The main components include:

- » **Application Data Modeling** provides automated procedures to discover sensitive columns and parent-child relationships; we then store the discovered results as data models that are reusable across multiple databases.
- » **Masking Format Library** provides comprehensive masking formats or templates for common types such as credit card numbers; it also includes functions or transformations to easily create reusable custom masking formats.
- » **Data Masking Definitions** assists in mapping masking formats to discovered sensitive columns creating reusable masking scripts; it also provides a workflow to mask data.
- » **Data Subsetting Definitions** helps to create reusable goal/condition based subsetting rules on a database; it also provides a workflow to generate subsets.

### **Q: How does the product preserve integrity of the data in an application?**

**A:** Oracle Data Masking and Subsetting employs two main techniques 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.

### **Q: Does the product include pre-configured masking formats or templates?**

**A:** Yes, Oracle Data Masking and Subsetting provides out-of-the-box masking formats or templates covering a broad range of sensitive data elements such as National identifiers of multiple countries, Credit card numbers of multiple vendors, Telephone numbers, Universal Product Codes and more.

### **Q: How does masking work with multi-byte or international characters?**

**A:** Several different masking operations support multi-byte or international characters such as UTF-8. The masking formats that can be used include: Array List, Shuffle, Substitute, Table Column, and User Defined Function.

### **Q: Which masking transformations or techniques are supported by the product?**

**A:** For simple masking use cases the product provides transformations such as fixed characters or numbers, column substitution, and SQL or regular expression based masking. The product also provides several advanced transformations such as:

- » **Shuffling** fields within a column in random fashion breaking one to one mapping between sensitive data elements. For example, columns containing salaries can be shuffled to break the employee-salary mapping.

- » **Blurring** or generating random numbers, characters, and dates within a user defined range.
- » **Encrypting** sensitive data using a key while preserving the format of the data. This transformation is useful when masked data sent to a third party has to be merged back along with further updates.
- » **Conditional masking** columns according to different conditions. For example, identifiers that belong to the United States can be masked using Social Security Number format and those that belong to the United Kingdom can be masked using National Insurance Number format.
- » **Compound masking** or grouping related columns together. For example, if you want to shuffle address fields like city, state, and country, then grouping city and the state will keep these columns together during the shuffling process.
- » **Deterministically** or consistently generating masked outputs for a given input across application schemas and databases. This makes it possible to mask names consistently or deterministically across different modules across your organization.

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

**A:** Each masking format or template has built-in logic to validate the generated masked output. For example, all credit card number masking formats perform LUHN checks on the masked credit card number. Also, the product provides an option to generate a preview to validate defined masking and subsetting criteria prior to the execution of masking and subsetting scripts.

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

**A:** Yes, the product supports migration of existing masking scripts for custom data types through user defined PL/SQL masking functions.

**Q: Which subsetting techniques are supported by the product?**

**A:** Oracle Data Masking and Subsetting simplifies the task of subsetting through its goal or condition based subsetting techniques. The goal can be a relative table-size such as extracting 1% subset of a table containing 10 billion rows for testing. Condition based subsetting can minimize risk exposure by discarding all user records created prior to a specific date, for example.

**Q: Do Oracle Data Masking and Subsetting work with applications such as Oracle E-Business Suite, Oracle Fusion Applications, Oracle PeopleSoft, and SAP?**

**A:** 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, 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 pre-packaged masking definitions for use with Oracle Data Masking and Subsetting.

**Q: Can I mask and subset databases running in Oracle Cloud?**

**A:** 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.



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

**A:** Yes, Oracle Data Masking and Subsetting can mask and subset data in DB2, Informix, SQLServer, Sybase, MySQL, and Teradata.

## Deployment and Administration

**Q: How do I download and install Oracle Data Masking and Subsetting?**

**A:** 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.

**Q: What are the different ways to mask and subset data?**

**A:** The product provides two modes to mask and subset data: In-Database and In-Export.

- » With In-Database masking and subsetting, the target production data first is copied (cloned) to a separate location. Then, Oracle Data Masking and Subsetting operates on the cloned data. After processing is complete, the resulting sanitized information is cloned to non-production databases.
  
- » With In-Export masking and subsetting, data is extracted from the target database, masking and subsetting rules are applied during the extraction process, 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.

**Q: What is the performance of the product?**

**A:** Oracle Data Masking and Subsetting delivers high speed masking and subsetting. High performance during masking and subsetting is achieved through the integration with Oracle Database Kernel and Data Pump module.

## More Information

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

**A:** For more information, please see the Oracle Data Masking and Subsetting page on Oracle Technology Network. A variety of helpful information is available online including data sheet, white paper, customer references, videos, tutorials, end-user documentation, virtual machine image, product blog, and a discussion forum.



**Oracle Corporation, World Headquarters**

500 Oracle Parkway  
Redwood Shores, CA 94065, USA

**Worldwide Inquiries**

Phone: +1.650.506.7000  
Fax: +1.650.506.7200

CONNECT WITH US

-  [blogs.oracle.com/oracle](https://blogs.oracle.com/oracle)
-  [facebook.com/oracle](https://facebook.com/oracle)
-  [twitter.com/oracle](https://twitter.com/oracle)
-  [oracle.com](https://oracle.com)

**Integrated Cloud Applications & Platform Services**

Copyright © 2016, Oracle and/or its affiliates. All rights reserved. This document is provided for information purposes only, and the contents hereof are subject to change without notice. This document is not warranted to be error-free, nor subject to any other warranties or conditions, whether expressed orally or implied in law, including implied warranties and conditions of merchantability or fitness for a particular purpose. We specifically disclaim any liability with respect to this document, and no contractual obligations are formed either directly or indirectly by this document. This document may not be reproduced or transmitted in any form or by any means, electronic or mechanical, for any purpose, without our prior written permission.

Oracle and Java are registered trademarks of Oracle and/or its affiliates. Other names may be trademarks of their respective owners.

Intel and Intel Xeon are trademarks or registered trademarks of Intel Corporation. All SPARC trademarks are used under license and are trademarks or registered trademarks of SPARC International, Inc. AMD, Opteron, the AMD logo, and the AMD Opteron logo are trademarks or registered trademarks of Advanced Micro Devices. UNIX is a registered trademark of The Open Group. 0615

