

# Oracle Database Migration Assistant for Unicode

## KEY FEATURES AND BENEFITS

THE MOST COMPLETE SOLUTION FOR MIGRATING DATABASES FROM LEGACY CHARACTER SETS TO THE UNICODE® STANDARD WITH MAXIMUM DATA INTEGRITY AND EFFICIENCY

### FEATURES

- Guided end-to-end migration workflow
- Intuitive graphical user interface
- Advanced data analysis and cleansing tools
- Migration profile support
- High-performance, scalable in-place conversion architecture
- Near-zero downtime migration model
- Comprehensive coverage of data types and database objects
- PL/SQL code scan to identify problem code patterns for Unicode migrations
- Progress monitoring
- Validation mode to check data integrity for compliance with the Unicode Standard
- Command-line scanner for automated problem data discovery

### BENEFITS

- Alleviate costly manual workload through automated migration tasks
- Greatly simplify data preparation process and prevent data loss
- Significantly reduce migration downtime

*Migrating databases to the Unicode® character set can be a daunting task even for the most seasoned DBAs. Without careful planning, accurate data analysis, choosing the best execution strategies, it may result in prolonged downtime, loss of data, or jeopardized system integrity. Oracle Database Migration Assistant for Unicode (DMU) is a next-generation migration tool that streamlines the entire migration process with an intuitive GUI that minimizes the DBA's manual workload and decision-making. It helps ensure all migration issues are addressed beforehand and the conversion of data is carried out correctly and efficiently.*

### About Oracle DMU

The DMU was first released in April 2011 on OTN as a free downloadable product. The latest DMU version 23.1 supports the migration of Oracle 23ai databases to Unicode and is the officially supported method for migrating Oracle databases to the Unicode character set. The DMU also supports migrating prior database releases of 11.2.0.4 and later. The legacy command-line utilities CSSCAN and CSALTER have been desupported starting from Oracle Database 12c.

### Guided End-to-end Migration Workflow

Migration to the Unicode character set is an intricate process that involves many different operational aspects which can be both time-consuming and resource intensive. Any misstep along the way can lead to data loss and serious business consequences. Oracle Database Migration Assistant for Unicode provides an end-to-end solution to the problem with a seamless workflow that drastically reduces administrative cost and complexity by covering:

- Enumeration – auto-identification of database objects containing textual data that requires conversion
- Scanning – comprehensive assessment of migration feasibility and discovery of potential data issues

- Address application impact of Unicode migrations
- Robust error-handling and failure recovery
- Health check on data integrity in databases encoded with the Unicode Standard
- Cleansing – sophisticated toolset for iterative data analysis and cleansing to ensure data safety
- Conversion – automated in-place data conversion to minimize time and space requirements

The workflow will guide the DBAs through the detailed migration process step-by-step and keep them informed of the real-time progress and outstanding tasks. It includes customization capabilities to allow the migration procedures to be performed in a manner most suited for the specific type of the database and runtime environment. Each component of the workflow is designed to achieve the highest possible migration accuracy and efficiency with the least amount of process overhead. Robust recovery and diagnostic mechanisms are built into the workflow to enable a fault-tolerant migration in the event of unforeseen runtime contingencies.

### **Easy-to-use Graphical Interface**

Oracle DMU comes with an intuitive graphical user interface that makes it remarkably easy to accomplish all the migration related tasks. The versatile interface is tightly integrated with the workflow to offer rich user interaction and visualization for effectively conveying migration information and status feedback. The otherwise challenging tasks such as identifying the cause of the data issues, applying data cleansing actions, converting the data, can all be done at the ease of a mouse click, keeping error-prone manual workload to a minimum.

### **In-depth Data Analysis**

As data is the most valuable asset of your organization, maintaining data integrity during the migration is of utmost importance. During the migration to the Unicode character set, data issues can arise due to different scenarios ranging from incorrect application configuration, to data expansion, storage of non-textual data in character columns, and more. Figuring out the underlying cause of the issue and taking corresponding cleansing actions to preserve the information is essential to a successful migration.

The DMU is capable of scanning all character data in the database to discover data convertibility issues that may compromise data integrity during the migration to the Unicode character set. The information about the data issues uncovered during the scanning phase can be accessed and saved via the scan report and the problem data report for further review and analysis outside of the tool. The DMU Command-line Scanner (DMU-CLS) can also be used to regularly monitor data convertibility issues and implement health check for Unicode databases.

The DMU Cleansing Editor greatly simplifies the process of investigating and cleansing problematic data. It provides unique visual features to systematically pinpoint the exceptional data and implement the best cleansing solutions. For databases containing mixed encoding data, the DMU can help determine the actual data encodings with the built-in character set detection feature and use them to migrate the data accordingly. The new extended VARCHAR2 type limit of 32767 bytes in

Oracle Database 12c guarantees that any VARCHAR2 column with the pre-12.1 limit of 4000 bytes or less can just be lengthened to accommodate longer values resulting from data expanding in conversion to Unicode. Neither truncation nor migration to the CLOB data type is now necessary for such columns. You can choose to perform a cleansing action immediately or schedule it to be deferred to the conversion downtime window if it is a schema-related change with application impact.

The DMU's Bulk Cleansing feature offers a convenient way to migrate multiple columns – for example, all columns in all tables of one or more application schemas – from byte length semantics to character length semantics. Migrating to character length semantics allows column sizes to be specified in terms of characters independent of the database character set, thereby effectively reducing the number of column length issues due to data expansion. You can also perform pattern-based cleansing to replace all occurrences of a source pattern with a target pattern, which can be very useful for resolving similar data convertibility issues that exist in multiple database objects.

The DMU Index Check feature enables the identification of indexes whose key sizes will exceed the maximum supported size as a result of the database conversion so that they can be analyzed and addressed before the migration downtime window.

The DMU also comes with migration profile support which allows you to export all migration-related settings from one database instance and import them into another database instance created as a clone of the original instance. This can be particularly useful for reusing and fine-tuning migration settings during the iterative trial migration process before performing the actual production migration.

### **Significantly Reduced Migration Downtime**

For most production systems, especially mission-critical applications, limiting the migration downtime to as short as possible is always a crucial consideration. Oracle DMU employs a novel in-place data migration strategy that focuses on only the data that needs to be converted. Since the majority of character data in real-world production databases is 7-bit ASCII, which does not change binary representation in the Unicode UTF-8 encoding, this strategy offers huge performance advantages over the traditional export/import approach, which is much more costly due to the processing of unnecessary data and the need to setup multiple instances. An innovative architecture leverages dedicated database server-side migration functionality and powerful parallel features to produce the maximum possible data throughput and scalability. It also utilizes built-in intelligence to evenly assign the workload among multiple worker threads and recommend the optimal execution plan and conversion method based on the data distribution characteristics. For large tables containing CLOB data, DMU provides the Convert using Data Pump method to speed up the conversion performance.

### Near-Zero Downtime Migration Model

The DMU offers a solution to achieve a near-zero downtime migration to the Unicode character set by integrating with the Oracle GoldenGate replication technology. You can set up a migration procedure that utilizes the DMU to perform the data preparation and in-place conversion while relying on Oracle GoldenGate to replicate any incremental data changes captured on the production system during the migration process. The DMU can generate the required Oracle GoldenGate configuration files to ensure that all incremental data is replicated correctly by taking into consideration any scheduled data cleansing actions already defined to address convertibility issues.

### Most Comprehensive Data Coverage

The DMU supports the migration of almost all Oracle data types that may directly or indirectly contain textual data. It also transparently deals with database objects, such as materialized views, indexes, constraints, and triggers, that are affected by conversion of tables so that they are properly synced up after the migration.

### Address Application Impact of Unicode Migrations

An important but often overlooked aspect of Unicode migration projects is the impact of the character set change on database applications. When moving from a non-Unicode character set to Unicode, especially from a single-byte character set to Unicode, there are various potential issues that must be considered and addressed beforehand to ensure that existing applications will continue to run correctly against the migrated Unicode database. Any changes in the database schema definitions, the maximum byte size of a character, the length semantics of character columns, and the behavior of functions dependent on the database character set, etc., could all require the application code to be updated accordingly to accommodate the new characteristics of the Unicode database.

The DMU offers the PL/SQL code scan feature to proactively scan PL/SQL applications to identify Unicode migration-related issues and common code patterns that may be problematic in the context of character set changes. Like the DMU data scan, the PL/SQL code scan feature will generate a scan report summarizing the findings along with the recommended actions, allow the user to interactively analyze the affected code modules, and export the problem code to spreadsheets for record keeping and sharing with application developers.

### Safeguard Data Integrity in Your Database

The DMU goes beyond facilitating migration to the Unicode character set by providing the capability to conduct ongoing health check of your post-migration database to maintain data compliance with the Unicode Standard. Even with a database that uses the Unicode character set, incorrectly configured applications may introduce invalid character codes into the database and cause data corruption. The DMU's Validation Mode feature can help expose the source of the issues and detect data problems, including presence of Unicode replacement characters (U+FFFD), ill-formed

Unicode byte sequences, non-shortest form UTF-8, irregular UTF-8 byte sequences (6-byte surrogate pairs), unpaired surrogates, and Unicode noncharacters, before these issues are even noted by end-users.

---

## 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](https://blogs.oracle.com)

 [facebook.com/oracle](https://facebook.com/oracle)

 [twitter.com/oracle](https://twitter.com/oracle)

Copyright © 2024, Oracle and/or its affiliates. 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, Java, MySQL, and NetSuite are registered trademarks of Oracle and/or its affiliates. Other names may be trademarks of their respective owners.