Accelerate/Streamline Your Unicode Migration: Oracle Database Migration Assistant for Unicode

Michael Yau, Senior Director
Weiran Zhang, Senior Development Manager
Server Globalization Technology
The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, and timing of any features or functionality described for Oracle’s products remains at the sole discretion of Oracle.
Agenda

- Unicode migration trend
- Migration process and common challenges
- Database Migration Assistant for Unicode
- Product demo
- Release plan
- Q & A
Unicode Migration Trend

• Rising demand for Unicode

Growth of Unicode on the Web

Source: Google Internal Data, compiled by Erik van der Poel
 Unicode Migration Trend (Cont.)

• Business drivers for migration
  – Global business expansion
  – Global IT infrastructure consolidation
  – Enterprise architecture standardization
  – Interoperability with popular technology stacks and packaged Apps

• Migrate earlier than later
  – Data volume could double every 2-3 years
  – Deferring the migration only makes it harder and more expensive later
Migrating a Database to Unicode

• It involves converting all text data (binary representation) in the database from the legacy character set encoding to the Unicode encoding
• It’s an expensive operation
  – Except for 7-bit ASCII data in UTF-8 (no change)
• It’s critical to preserve data integrity
Unicode Migration End-to-End Steps

Project Planning
- Confirm business requirements
- Determine project timeline
- Identify business and technical constraints
- Allocate machine and human resources

Data Preparation
- Assess conversion feasibility
- Analyze data issues and manually resolve them by writing DBA scripts and using 3rd-party tools
- Require multiple iterations and rehearsals

Data Conversion
- Schedule system downtime for production migration
- Perform backup
- Set up separate Unicode target instance
- Move data with EXP/IMP or Data Pump

Post-Conversion Tasks
- Application code sync-up
- Data validation testing
- Application integration testing
- User acceptance testing

Go Live
Unicode Migration Challenges

• Limited down-time window
• Data integrity
  – Invalid data (pass-through)
  – Data expansion
  – Binary data stored in textual columns
• Resource and cost constraints
• Dependent objects (index, constraint, trigger, partitioned range, materialized view, etc.)
• Application code impact
• Incremental data changes before migration
• Failure recovery to restore database consistency
• Lack of domain expertise
No Existing End-to-End Migration Solution

• CSSCAN – character set scanner
  – Identify all textual data
  – Analyze migration feasibility
  – Generate plain-text report

• CSALTER script
  – Updates metadata in the data dictionary
  – Only applicable if no data requires conversion

• EXP/IMP and Data Pump
  – Time and resource intensive
  – May convert much more data than necessary (typically < 1% of data requires conversion)
Introducing Oracle Database Migration Assistant for Unicode (DMU)
DMU Key Features

- Streamlined end-to-end migration workflow
- Intuitive GUI interface with data visualizations
- Automated migration tasks (Scan/Cleanse/Convert)
- Advanced data analysis and cleansing tools
- Inline conversion approach to drastically reduce migration time and resource requirements
- Maximize performance with scalable server-side architecture
DMU Key Features (Cont.)

• Comprehensive coverage of Oracle datatypes and database objects
  • char, varchar2, long, clob, adt, nested table, varray, XMLType, anydata, anydataset, …

• Customizable workflow and execution plan
• Real-time progress monitoring
• Recovery and diagnostics for better fault-tolerance
• Health check for Unicode data integrity
DMU Inline Migration Architecture

- GUI Guided End-to-End Workflow
- Migration Logic (Scan/Cleanse/In-Place Convert)
- Parallel Processing and Load Balancing
- Post-Conversion Operations
- Unicode Health Check
- Status Monitoring

DMU

Oracle DB

JDBC Connection

DMU Repository

Scan

Convert

PQ

CTAS

UPDATE

METHOD

METHOD

Trusted Callouts

Custom SQL functions
DMU Key Benefits

• Significantly reduce migration time/costs and protect data integrity
  – Simplify the migration process
  – Alleviate DBA manual workload with automated migration tasks
  – Significantly speed up the process of identifying and cleansing problem data
  – Significantly reduce downtime with In-Place Conversion technology vs. export/import method
  – Error-handling and failure recovery
  – Lessen the bottleneck of requiring DBA with in-depth character set migration expertise
• Post-migration: Health check on Unicode data integrity and identify data path and application support issues
DMU PRODUCT DEMO

(Also visit our demo in Moscone West Booth W-085)
Feedback from a DMU Beta Customer  
- Migrating from US7ASCII to AL32UTF8

We have been impressed with the **speed, flexibility and easy-of-use** of the DMU tool. Although we have relatively small databases, the DMU tool has scanned them consistently faster than the csscan tool, with our largest database (124GB) taking **one-tenth of the time**. We are converting our databases from the US7ASCII character set to AL32UTF8, but we have 8-bit latin-1 characters throughout these databases. The 'Database Scan Report' option provided a **simple, but flexible summary of the data issues** in the database, clearly highlighting the problematic data.

In some databases there were characters that were not represented in the WE8MSWIN1252 character set. The DMU tool provided an easy-to-use 'data cleansing' tool to alter or remove this data. We would have appreciated the ability to 'globally replace' these invalid characters in a table column with null or another character.
Feedback from a DMU Beta Customer (Cont.)

As part of the DMU data conversion process we specified WE8MSWIN1252 as the 'assumed character set' for any characters outside the ASCII range, thereby enabling us to successfully convert these databases directly from US7ASCII to AL32UTF8 in a single execution. The duration of the conversion process varied widely, depending on the volume of convertible data and CLOB data in the database, but most conversions were completed well within 90 minutes.

We have found the export/import method of converting databases from US7ASCII to AL32UTF8 both cumbersome and time-consuming. Prior to testing the DMU tool we attempted to convert copies of two Production databases using this traditional method...Issues arose when we attempted to import the data into AL32UTF8 databases, mostly related to insufficient storage in tablespaces. These problems required multiple attempts to import the data successfully.

With over 100 databases yet to be converted we anticipate that the extra time required with export/import method will lengthen our project timetable and add risk to our Production implementation over a 3-day weekend. In contrast, the DMU tool converted these same databases from US7ASCII to AL32UTF8 'in-place' within 90 minutes. The DMU tool will add robustness and confidence to the database conversion process and our project planning. We are keenly awaiting the production release of the DMU tool to save us from the otherwise huge effort to complete this transition.”
DMU Release Plan

• Standalone Product
  – Distributed on OTN
  – NO COST! (with paid Database license)

• Oracle Support
  – Supported with Database Support License

• General Availability
  - Planned for release in the next 12 months (Currently in Beta)
  - Watch out for OTN news announcement!

• Supported Database Releases & Platforms
  – 10.2.0.4.4, 10.2.0.5.1, 11.1.0.7.5, 11.2.0.1.3
  – Linux64, Solaris64, HPUX64, AIX

• Email questions to infonls_migration_us@oracle.com
Questions?
SOFTWARE. HARDWARE. COMPLETE.