Near-Zero Downtime Database Migration to Unicode – Overview

Table of Contents

.

Executive Overview

Companies that do business in the global market require the use of a universal character set in order to support multilingual users and organizations. Unicode is a universal character set encoding that enables information from any language to be stored using a single character set. Unicode provides a unique code value for every character, regardless of the platform, program, or language. Unicode encoding is becoming the database character set encoding of choice. However, there are many databases still using legacy character sets, as companies have concerns over the complexity of the Unicode migration process. It is perceived to be challenging together with the downtime window requirement which have often prevented them from migrating.

In this paper, we will review a near-zero downtime Unicode migration process using Oracle Database Migration Assistant for Unicode, Oracle GoldenGate and Oracle Recovery Manager, by following the logical workflow of the Oracle Database Migration Assistant for Unicode to ensure that the migration procedure is performed correctly and efficiently.

Introduction

Migrating databases to Unicode can be a daunting task. When performed without careful planning, accurate data analysis, choosing the best execution strategies, it may result in prolonged downtime, loss of data, or jeopardized system integrity. And when companies decide to go forward with their database migration they also face challenges with IT continuity, as users expect critical applications to be continuously available. Oracle Database Migration Assistant for Unicode is a tool that can assist customers to minimize the migration efforts. Oracle GoldenGate is a software product for real-time data integration and replication which is used to implement a near-zero downtime migration.

A Near-Zero Downtime Database Migration to Unicode is discussed in two white papers:

» Near-Zero Downtime Database Migration to Unicode - Overview (this paper).

   This paper provides high level information for the near-zero downtime migration process.

» Near-Zero Downtime Database Migration to Unicode - Procedures.

   This paper provides detailed information for the near-zero downtime migration process and setup procedures.

This white paper is divided into the following sections:

» Character Set Migration Considerations

   We review the character set migration process and potential issues to consider.

» Brief overview of the products used in the near-zero downtime Unicode migration:

   » Oracle Database Migration Assistant for Unicode
   » Oracle GoldenGate
   » Oracle Recovery Manager

» Near-Zero Downtime Unicode Migration Process

   We describe the Unicode migration process using the products mentioned above.

» Generating Near-Zero Downtime Migration Configuration Files

   We review the configuration file generation, and the configuration and extract types that the Oracle Database Migration Assistant for Unicode generates.

The following sections list the minimum software requirements:

» Oracle Database version 11*g* or later
» Oracle Database Migration Assistant for Unicode (DMU) version 2.1 or later
» Oracle GoldenGate (OGG) version 12.1.2.1.0 or later

Character Set Migration Considerations

A database defined in a legacy character set can store data only in one or a few languages as determined by its character set. If a need arises to store data in other languages in this database, the character set of the database must be changed to a superset (e.g. character set A is a superset of character set B if A supports all characters that B supports) that defines both the existing and the new required characters. If the identified superset is also a binary superset (e.g. character set A is a binary superset of character set B if A supports all characters that B supports and all these characters have the same binary representation in A and B.), the change of the database character set declaration must happen only in metadata (schema information) in the data dictionary. All existing data, per definition, remains valid in the binary superset. The change is simple and fast. If the identified superset is not a binary superset, binary character codes (byte representation) must be converted from the representation in the old character set to the representation in the new character set that is, re-encoded.

An important observation, relevant to determining if existing data requires conversion, is that even if the old database character set is not a binary subset of the new database character set, many characters may actually have the same binary codes in both character sets. In particular, all standard ASCII characters with codes between 0 and 127 form a binary subset of each Oracle database character sets supported on ASCII-based platforms. Therefore, even if the planned database character set change is, for example, from WE8ISO8859P1 to AL32UTF8 - which are not in a binary subset-superset relationship - but all characters that are in the database have codes in the range 0 to 127, then no characters really need to be converted, because all characters will have the same codes after the change.

If character data in a database to be converted can be classified into data that needs no conversion, called **changeless**, and data that requires conversion, called **convertible**, a lot or resources can be saved by skipping the conversion of the changeless data. When the character codes are converted, various issues may arise. The most common one is that the new representation may have more bytes causing the converted value to no longer satisfy the containing column's length constraint or even the maximum length of its data type. For conversion to succeed, those issues have to be identified and cleansed first, for example, by lengthening columns, shortening texts, or migrating to a larger data type.

The process of identifying the suitable superset, classifying character data as changeless or convertible, checking for issues and conversion requirements, fixing the issues, changing the declaration in the data dictionary and converting data as needed is called character set migration. The issues and the subset/superset relationship of existing data in the database may be identified with the help of the Oracle Database Migration Assistant for Unicode.

The following sections are some of the migration potential issues to consider:

» Data Expansion

When you migrate your database from a legacy encoding to Unicode, character values will likely expand in conversion because their encodings will have more bytes and hence increase space requirements for the database. A further issue is that the widths for CHAR and VARCHAR2 columns may not be sufficient after the character set has been migrated to Unicode. Thus, there is a risk of the data being truncated. The column

length constraints have to be increased, or, if they are already at the data type limit, the columns might need to be migrated to a larger data type.

» Invalid Binary Storage Representation of Data

For user data, a common problem is that the data in a column is not actually in the declared database character set. Instead, it is in another character set or it is binary or perhaps there are multiple character sets in a single column.

» Downtime Window

The downtime window is the time when all applications accessing the production database are shut down and the database is accessible only to the migration tools. A downtime window is required to perform the actual conversion of the database.

» Failure Recovery

When you perform a migration, the process may abruptly terminate for many reasons, such as a software defect or hardware failure. In that case, the database could be left in an inconsistent state. If the migration utility is not able to resume the migration, you may need to bring your database back to a consistent state. Dealing with failures during the migration process is a necessary part of your migration planning.

» Application Impact

The character set migration of a database, especially from a single-byte to a multi-byte character set, may have a significant impact on applications using this database.

Oracle Database Migration Assistant for Unicode

The Oracle Database Migration Assistant for Unicode (DMU) is an intuitive user-friendly GUI tool that helps you streamline the migration process through an interface that minimizes the workload and ensures that all migration issues are addressed, along with guaranteeing that the data conversion is carried out correctly and efficiently.

Some of the advantages that DMU provides:

» Guides you through the workflow
It offers a logical workflow to guide you through the entire process of migrating character sets.
   » Enumeration - Identify database objects containing data that requires conversion.
   » Scanning - Scan its contents for convertibility issues, issues that may cause data loss or corruption.
   » Cleansing - Data issues that have been identified in the scanning process must be resolved before the database can be converted.
   » Conversion - Convert the database to the target Unicode character set with in-place data conversion to minimize time and space requirements.
» Offers suggestions for handling problems
The DMU can help you when you run into certain problems, such as errors or failures during the scanning or cleansing of the data.

» Offers progress monitoring
The DMU provides a GUI to visualize how the steps are progressing.

» Offers interactive visualization features
The DMU enables you to analyze data and see the results in the GUI in an interactive way. It also enables you to see the data itself in the GUI and cleanse it interactively from identified migration issues.

» Provides the only supported tool for inline conversion

With the DMU, Oracle Database supports inline conversion of database contents, which allows the updates of data that requires conversion in the same database instance. This offers performance and security advantage over other existing conversion methods.

» Allows cleansing actions to be scheduled for later execution during the conversion step

Postponing of cleansing actions, such as data type migration, ensures that the production database and applications are not affected until the actual migration downtime window.

» Generates GoldenGate configuration files for a near-zero downtime Unicode migration

DMU can generate configuration files for the Oracle GoldenGate which can be used to setup an environment for a near-zero downtime Unicode migration procedure. The generated parameter and script files are used by the Oracle GoldenGate Extract and Replicat processes for replicating data. The generated configuration files also adhere to the scheduled column modification and assumed character set tagging performed in DMU.

Oracle GoldenGate

Oracle GoldenGate (OGG) is real-time data integration software that provides real-time changed data capture, routing, transformation, and delivery across heterogeneous databases. It captures transactions from a source database by reading online transaction logs, transforming the data when needed, and applying those transactions with guaranteed integrity to a target database.

Oracle GoldenGate consists of 4 distinct modules and components:

» Oracle GoldenGate Manager
» Oracle GoldenGate Extract
» Oracle GoldenGate Trail files
» Oracle GoldenGate Replicat

Oracle GoldenGate Manager

Manager is the control process of Oracle GoldenGate. Manager must be running on each system in the Oracle GoldenGate configuration before Extract or Replicat can be started, and Manager must remain running while those processes are running so that resource management functions are performed.

Oracle GoldenGate Extract

The Extract process runs on the source system and is the extraction (capture) mechanism of Oracle GoldenGate.

You can configure Extract in one of the following ways:

» Initial load: Extract extracts (captures) a current, static set of data directly from their source objects.
» Change synchronization: To keep source data synchronized with another set of data, Extract captures DML and DDL operations after the initial synchronization has taken place.

Oracle GoldenGate Trail Files

To support the continuous extraction and replication of database changes, Oracle GoldenGate stores records of the captured changes temporarily on disk in a series of files called a trail. A trail can exist on the source system, an intermediary system, the target system, or any combination of those systems, depending on how you configure Oracle GoldenGate. On the local system it is known as an extract trail (or local trail). On a remote system it is known as a remote trail.

Oracle GoldenGate Delivery (Replicat)

The Replicat process runs on the target system, reads the trail on that system, and then reconstructs the DML or DDL operations and applies them to the target database.

You can configure Replicat in one of the following ways:

» Initial load: Replicat can apply a static data copy to target objects or route it to a high-speed bulk-load utility.
» Change synchronization: When configured for change synchronization, Replicat applies the replicated source operations to the target objects using a native database interface or ODBC, depending on the database type. To preserve data integrity, Replicat applies the replicated operations in the same order as they were committed to the source database.

Oracle Recovery Manager

Oracle Recovery Manager (RMAN) is a database tool that fully integrates with the Oracle Database for managing the process of making backups, restoring and recovering from these backups. In the process described in this document, it is used for backing up and cloning the source database. The cloned database is used as an Initial Load on the target system in the Oracle GoldenGate (OGG) configuration.

Near-Zero Downtime Unicode Migration Process

The diagram below depicts a near-zero downtime migration process using the following products:

» Oracle Database Migration Assistant for Unicode (DMU)
    » scan database objects for potential data issues
    » cleanse data to ensure data integrity
    » generate Oracle GoldenGate configuration files
    » convert the database to Unicode
» Oracle Recovery Manager (RMAN)

    RMAN is used for backing up and cloning the source database. The cloned/target database is used as an Initial Load on the target system in the Oracle GoldenGate configuration.
» Oracle GoldenGate (OGG)

    OGG is used for replicating data - change synchronization from the source database to the target database.

Initial Load/Clone

```
                    ┌─────────────────┐
                    │    (5) RMAN      │
                    │ Backup and Clone │
                    └─────────────────┘
```
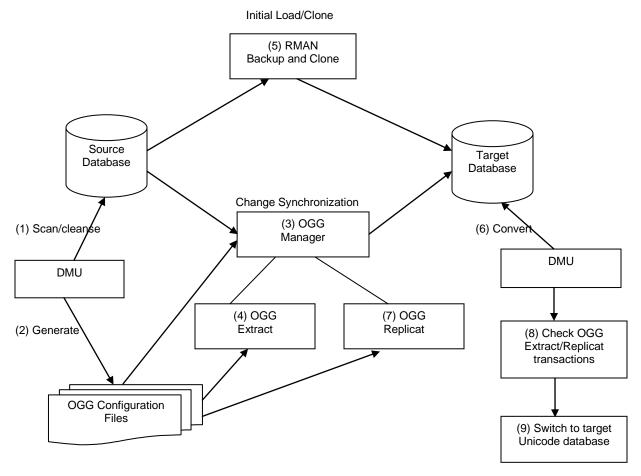

Fig. 1 – Near-Zero Downtime Unicode Migration Process

The following section provides the high level migration process for a near–zero downtime database migration to Unicode:

1. Scan and cleanse textual data for convertibility issues using DMU until the database is ready for conversion.

2. Generate Oracle GoldenGate configuration files using DMU, these configurations files are needed by the Oracle GoldenGate processes.

3. Start the Oracle GoldenGate Manager process, the Manager process is the parent process responsible for managing its processes (Extract and Replicat) and files.

4. Start the Oracle GoldenGate Extract process to start capturing committed transactions from the source database.

5. Backup and capture the checkpoint SCN value, and clone the source database using RMAN. The cloned database is used as the target database as an Initial Load.

6. Convert the target database from the legacy character set to Unicode using DMU.

7. Start the Oracle GoldenGate Replicat process (by providing the checkpoint SCN value captured during the backup step above) for applying the committed transactions to the target database.

8. Check that all the transactions have been extracted and replicated by OGG.

9. Switch users and applications to the target Unicode database.

The more detailed migration process is described in the accompanied white paper "Near-Zero Downtime Database Migration to Unicode – Procedures".

Generating Near-Zero Downtime Migration Configuration Files

To correctly replicate the incremental data - change synchronization on the target Unicode database, Oracle GoldenGate requires certain parameter files containing the information about the source database schemas, tables, columns, character sets, and any required data cleansing actions (e.g. lengthening of columns, migration to a larger data type, and assumed character set tagging) to address the convertibility issues to Unicode. As DMU repository already contains this information, DMU provides the functionality of exporting this information in the form of the parameter files for OGG Manager, Extract and Replicat processes and script files (called as Obey scripts) for starting and stopping data replication.

DMU generates OGG parameter and script files using:

» Classic capture mode (database version 11.2.0.3 or older)
» Integrated capture mode (database version 11.2.0.4 or newer)

For Oracle Database 12*c*, DMU can generate OGG parameter and script files for container databases (CDBs) and non-container databases (Non-CDBs).

There are two types of configurations that DMU generates:

» Local Replicat – OGG Extract and Replicat processes run on the same host.
» Remote Replicat – OGG Extract and Replicat processes run on different hosts.

More information about the generated configuration files is available in the accompanied white paper "Near-Zero Downtime Database Migration to Unicode – Procedures".

Conclusion

For mission-critical applications, users expect systems to be continuously available 24/7. Using Oracle Database Migration Assistant for Unicode (DMU) together with Oracle GoldenGate (OGG), a database migration to Unicode can be achieved in a near-zero downtime window with minimized migration efforts.

Oracle Database Migration Assistant for Unicode (DMU) provides the following benefits:

» Identify database objects containing textual data that require conversion.
» Scan for potential data issues
» Cleanse to ensure data safety
» Generate parameter and script files for Oracle GoldenGate configuration
» Convert the database to Unicode

Benefits from using Oracle Database Migration Assistant for Unicode (DMU) with Oracle GoldenGate (OGG)

DMU can generate the parameter and script files for Oracle GoldenGate which can be used to setup a near-zero downtime Unicode migration procedure. The generated parameter and script files are needed for the Oracle GoldenGate processes in order to replicate incremental data changes correctly from the source database to the target database. The configuration file generation honors the scheduled column modifications and assumed character set tagging defined in DMU.

For further information, see the links below:

» Near-Zero Downtime Database Migration to Unicode – Procedures
» Oracle Database Migration Assistant for Unicode Documentation
» Oracle GoldenGate Documentation