

ORACLE DATABASE 12C

グローバル化・サポート - 新機能

主要な機能と利点

新機能

- Unicode 6.1 をサポート
- Unicode 照合アルゴリズム準拠による照合サポート
- 複数の新しいロケールに対応
- Database Migration Assistant for Unicode

利点

- 最新の Unicode 6.1 の文字定義に対応
- 柔軟な機能性にて業界標準の多言語照合に対応
- ロケール対象範囲の拡大により、アプリケーション・ローカライゼーションのサポート対象も拡大
- 停止時間の大幅削減、コストの抑制、データベースの Unicode キャラクタ・セットへの移行作業の簡略化を実現

Oracle Database 12c では、Unicode® キャラクタ・セットでのデータベースの構築や、標準準拠の多言語エンタープライズ・アプリケーションの開発を支援する一連の新機能を採用し、グローバル化・サポート力を充実させました。オラクルでは、あらゆる言語で文字データの格納、処理、取得が可能な、Unicode ベースのシステム・アーキテクチャを推奨しています。この新しいデータベース機能により、ビジネス要件に沿った完全な Unicode ソリューションの構築に、高い利便性と業界内の互換性が提供されます。

Unicode 6.1 をサポート

Unicode 規格では汎用的なキャラクタ・セットを定義し、世界の表記法の大半で使用されている文字をエンコードできるようにしています。これにより、プラットフォームやプログラミング言語から独立して、文字情報を統一的に表現できます。オラクルでは、Oracle 7 よりこの Unicode キャラクタ・セットをサポートしていましたが、Oracle Database 12c では、このサポート範囲を Unicode 規格バージョン 6.1 にまで拡大しました。

このバージョン 6.1 は 2012 年 1 月にユニコード・コンソーシアムよりリリースされたものです。これはすべての前バージョンの上位となるもので、中国やその他のアジア諸国、アフリカ諸国の文字が新たに Unicode 文字データベースに追加されています。バージョン 6.1 でエンコード可能な文字は 11 万超に及びます。さらに文字属性と照合アルゴリズムも改良されています。

オラクルの Unicode データベース・キャラクタ・セット AL32UTF8 と各国語キャラクタ・セット AL16UTF16 の定義も、Unicode 6.1 に合わせてアップデートしました。これらのキャラクタ・セットを Oracle Database 12c で使用することで、もっとも包括的な文字レパートリーにて、多言語データベースのあらゆる文字データ処理ニーズに対処できるようにしています。

Unicode 照合アルゴリズム (UCA) 準拠

言語が異なれば、文字列の比較やソートを実行する際のルールも異なります。また文字列の順序は一般的に、その言語のバイナリ表現の順序とは異なります。アプリケーションのグローバル化を成功裏に導くには、情報の比較や表示において、エンドユーザーの言語的慣習に沿うことが重要になります。

Unicode 照合アルゴリズム (UCA) は、Unicode 文字列の語順を定義する Unicode 規格です。UCA では、すべての Unicode 文字を合理的にデフォルト照合するための、デフォルト Unicode 照合基本テーブル (DUCET) が定義されています。DUCET はカスタマイズが可能であるため、特定の言語だけの特殊な語順にも対応できます。UCA は国際標準である ISO14651 と完全な互換性を持ちながらも、拡張機能により照合作業の柔軟性を確保しています。

Oracle Database 12c では、既存の単一言語照合機能および多言語照合機能に加え、UCA 準拠の言語照合を導入し、これにより Unicode 規格バージョン 6.1 への適合を実現しています。おもな機能は以下のとおりです。

- Unicode 6.1 DUCET に基づいた完全な照合順序
- 4 つの照合レベルまでの多段階比較アルゴリズム
- 可変文字（空白文字、句読点、特定の記号）に重み付けをしてソートできる構成可能なオプション
 - Blanked
 - Non-ignorable
 - Shifted
- スペイン語、伝統的スペイン語、フランス語（カナダ）、デンマーク語、タイ語、中国語簡体字（ピンイン、画数、部首）、中国語繁体字（画数、部首）、日本語、韓国語など、言語に合わせた 12 種類の UCA 照合

UCA は、多言語データのソートに最適なメカニズムとして推奨されています。この新たな UCA 照合により、多言語データの検索、ソート、マッチングをより詳細にコントロールできるようになりました。

また言語的な操作においては、相対的な順序を評価する前に文字データを照合キーと呼ばれるバイナリ値に変換します。この照合キーは、Oracle では RAW データ型で表現されるため、VARCHAR2、NVARCHAR2、および RAW の各データ型の最大長が 32,767 バイトまで拡張されたことを受け、Oracle Database 12c では長いテキストを高い精度でソートできるようになりました。

新しいロケール対象範囲

グローバリゼーション・サポートの対象範囲を拡大し、急速に進化する顧客要件に対応する継続的な取組みの一環として、Oracle Database 12c では、サポートするデータベース・ロケールに新たに 12 の言語と 32 の地域を加え、アジア、アフリカ、南北アメリカ、ヨーロッパの対象範囲を拡大しました。

- 新しい言語 – アムハラ語、アルメニア語、ダリー語、ディベヒ語、クメール語、ラオ語、ボスニア語（ラテン文字）、マルタ語、ネパール語、ペルシャ語、シンハラ語、スワヒリ語
- 新しい地域 – アフガニスタン、アルメニア、バハマ、ベリーズ、バミューダ、ボリビア、ボスニア・ヘルツェゴビナ、カンボジア、カメルーン、コンゴ・ブラザビル、コンゴ民主共和国、エチオピア、ガボン、ホンジュラス、イラン、コートジボワール、ケニア、ラオス、モルディブ、マルタ、モンテネグロ、ネパール、ナイジェリア、パキスタン、パラグアイ、セネガル、セルビア、スリランカ、タンザニア、ウガンダ、ウルグアイ、ザンビア

さらに、コプト暦に基づく暦法で、日数が 5 日または 6 日の 13 月がある、エチオピア暦もサポート対象に追加しました。

Oracle Database Migration Assistant for Unicode (Oracle DMU)

Unicode キャラクタ・セットへの移行は複雑なプロセスで、多くの時間とリソースをかけることにもなりかねない、さまざまな処理を伴います。またプロセスに間違いがあれば、データ損失の恐れもあり、ビジネスに影響が及ぶことも考えられます。Oracle Database Migration Assistant for Unicode は、直感的な GUI によって全体的な移行プロセスを効率化し、DBA の手作業や意思決定を最小限に留める次世代の移行ツールです。また、移行に伴うすべての問題にあらかじめ対処し、正確かつ効率的にデータ移行を実施するうえでも効果的です。Oracle DMU の移行ワークフローは次のとおりです。

- 特定 - 変換が必要なテキスト・データを含むデータベース・オブジェクトを自動的に特定します。
- スキャン - 移行の実現可能性を総合的に評価し、データに関する潜在的な問題を見つけ出します。
- クレンジング - 反復データ分析およびクレンジングが可能な高度なツールセットにてデータの安全性を確保します。
- 変換 - データのインプレース変換を自動実行することにより、時間的および領域的要件を最小限に押さえます。

Oracle DMU は無料ダウンロードが可能な製品として、2011 年 4 月に OTN にて初めてリリースされました。最新の Oracle DMU はバージョン 1.2 で、Oracle Database 12c にバンドルされており、データベースを Unicode キャラクタ・セットに移行する方法として公式サポートされています。また Oracle DMU は、以前の Oracle Database リリースである 10.2、11.1、11.2 の一部の移行もサポートしています。以前のコマンドライン・ユーティリティである CSSCAN と CSALTER はサポート対象外となりました。

Unicode キャラクタ・セットへの移行においては、アプリケーションにおける不正確なキャラクタ・セットの構成、列の上限を超えたデータの拡大、文字の列中の非テキスト・データなど、さまざまな理由でデータの問題が発生する可能性があります。Oracle DMU では、ツールやサポートを提供することで、そうしたデータ問題を診断し、適切なクレンジング・ソリューションを適用できるようにしています。さまざまなエンコーディング・データを含むデータベースにおいては、Oracle DMU を活用して実際のエンコーディングを特定し、それに基づいてデータを移行します。Oracle Database 12c では、VARCHAR2 型の上限が 32,767 バイトに拡張されたため、上限が 4,000 バイトであった 12.1 以前の VARCHAR2 の列も拡張が可能となり、Unicode への変換に伴うデータ拡大で長くなった値にも対応可能となりました。そのためこうした列には、切り捨てや CLOB データ型への移行を行う必要もありません。クレンジングはすぐに実行することもできますし、アプリケーションへの影響を伴うスキーマ関連の変換作業の場合は、その変換の停止時間中に実行するよう設定することも可能です。

移行に伴う停止時間を実質的に短縮するため、Oracle DMU では、変換すべきデータだけにフォーカスする高度なインプレース移行戦略を採用しています。従来のインポート/エクスポート手法では、必要のないデータを処理したり、複数のインスタンスを設定する必要があったりしたため、コストがかかっていました。しかし実際のデータベースでは文字データの大半は 7 ビット ASCII であり、Unicode UTF-8 エンコーディングにてバイナリ表現を変換しないため、この戦略にはそうした従来の手法を上回る大きなパフォーマンス上のメリットがあります。画期的な新しいアーキテクチャでは、専門特化したデータベース・サーバー側移行機能と便利なパラレル機能により、データのスループットとスケラビリティを最大化しています。またインテリジェント機能を組み込んだことで、複数のワーカー・スレッド間でのワークロードの均等割当てが可能になったほか、データ・ディストリビューションの特性に基づいた変換方法や最善の実行プランも推奨されるようになっていきます。

Oracle DMU は Unicode キャラクタ・セットへの移行を支援するだけでなく、移行後のデータベースを継続的にチェックする機能により、データの Unicode 規格への準拠を常に確保します。Unicode キャラクタ・セットを使用するデータベースでも、アプリケーションの構成に誤りがあると、無効な文字コードがデータベースに取り込まれ、データ破損を招くことがあります。Oracle DMU の検証モード機能は、こうした問題の発生源を明らかにし、エンドユーザーが気づく前にデータの問題を検出するうえで効果的です。

お問い合わせ先

Oracle Database 12c のグローバル化・サポートについて、詳しくは oracle.com を参照するか、+1.800.ORACLE1 でオラクルの担当者にお問い合わせください。



Copyright © 2009, Oracle and/or its affiliates. All rights reserved.

本文書は情報提供のみを目的として提供されており、ここに記載される内容は予告なく変更されることがあります。本文書は一切間違いないことを保証するものではなく、さらに、口述による明示または法律による黙示を問わず、特定の目的に対する商品性もしくは適合性についての黙示的な保証を含み、いかなる他の保証や条件も提供するものではありません。オラクルは本文書に関するいかなる法的責任も明確に否認し、本文書によって直接的または間接的に確立される契約義務はないものとします。本文書はオラクルの書面による許可を前もって得ることなく、いかなる目的のためにも、電子または印刷を含むいかなる形式や手段によっても再作成または送信することはできません。

Oracle は米国 Oracle Corporation およびその子会社、関連会社の登録商標です。その他の名称はそれぞれの会社の商標です。0109