Oracle Database 10*g*: グローバリゼーション・テクノロジの有効性

オラクル・ホワイト・ペーパー
2005 年9 月

ORACLE

Oracle Database 10*g*:
グローバリゼーション・テクノロジの有効性

Oracle Database 10*g*:
グローバリゼーション・テクノロジの有効性

グローバリゼーション・サポート - 概要

今日の企業は、ビジネスの最適化や飛躍的な拡大を目論む前例のないチャンスに恵まれています。WWW（World Wide Web）の登場により、インターネットやイントラネット・アプリケーションは、世界規模で公開される可能性を持つようになりました。世界人口の 92%が英語以外の言語を話し、国際的なインターネット・コミュニティが急速に拡大する中、チャンスは目前にあると言えます。統計的には、コンテンツが母国語で表示される場合、消費者が購買する可能性は 4 倍になると言われています。したがって、グローバルな市場で成功するためには、多言語に対応するアプリケーションが必要です。

Oracle は、どのような価値を提供できるでしょうか。第一に、ユーザーは、あらゆる言語でデータを保存、取得、更新できる必要があります。Oracle は、多言語標準である Unicode 4.0 に完全に対応しています。UTF-8、UTF-16 への対応により、現在世界で使用されているほとんどの言語を簡単にエンコードすることができます。これにより、単一のデータベース内で、またはグリッドの一部として、多言語を開発、配置、ホスティングできます。また、Oracle が提供する柔軟性により、すべてのデータを Unicode データベースに保存したり、選択した列を Unicode データ型として保存したりすることができます。もう 1 つの重要な機能としては、母国語でローカライズされた情報の表示があります。日時、通貨記号、デリミタ、照合順などは、Oracle でシームレスに処理されます。Oracle のローカライゼーション・サポートはほぼ包括的ですが、グラフィカル・ツールである Oracle Locale Builder ユーティリティを使用することで、特別なカスタマイズを行うこともできます。

通常、ほとんどの Web ベースのアプリケーションは複数層です。Oracle は、Unicode データの格納や取り出しを行うための様々なデータベース・アクセス製品を提供します。Oracle は、Java や C/C++など、最も一般的に使用されるプログラミング言語をサポートします。データベースとクライアント・プログラム間でデータを透過的に変換して、クライアント・プログラムがデータベースのキャラクタ・セットに依存しないようにします。

これらすべてを統合するには、状況に応じてレガシー・データやアプリケーションを Unicode 環境に移行することが必要になります。Oracle は、Character Set Scanner ユーティリティを備えているため、データの損失や切捨てなど、移行に関連する問題やデータベースが全体に与える影響の分析を事前に行うことができます。また、Oracle は、Globalization Development Kit（GDK）を提供しています。このキットは、開発プロセスを簡素化し、グローバル環境の対応したインターネット（Java および PL/SQL）アプリケーションの開発コストを削減します。

Oracle Corporation **発行**「Oracle Database 10*g*: The Power of Globalization Technology」**の翻訳版です。**

グローバリゼーション・テクノロジの有効性／はじめに

グローバリゼーションは、ビジネスを他の国々に拡大することを可能にしてくれます。世界各国の言語や商慣習に対応したアプリケーションやソフトウェアを作成することは、容易な作業ではありません。これには、単なる HTML コンテンツやメッセージの翻訳よりもはるかに多くのことが要求されます。

National Language Support、その他のグローバリゼーション・サポート

Oracle のほとんどの顧客は、National Language Support（NLS）という用語を理解し、その使用経験も豊富です。Oracle の特定のリリースを注文する顧客は全員、同じバージョン、同じバイナリ・ファイルを入手します。それにもかかわらず、Oracle データベースは世界中で販売されています。どうして、このようなことが可能なのでしょうか。National Language Support により、ユーザーはデータを各国の言語およびロケールで保存、処理、取得できます。NLS は、データベース・ユーティリティ、エラー・メッセージ、ソート順、日時、通貨、数字、カレンダの表記規則が、自動的にユーザーの言語およびロケールに対応することを保証しています。Oracle が、各データベース・インスタンスに対して、1 つの言語を単に処理する以上の機能を提供する現在、National Language Support という用語は時代遅れかもしれません。

組込みユニバーサル・サポート

Oracle の National Language Support アーキテクチャは、Oracle NLS Runtime Library（NLSRTL）により実装されています。NLS Runtime Library は、顧客には公開されませんが、SQL、PL/SQL などで提供される一連の包括的な API です。これにより、正しいテキストおよび文字の処理、文化的、言語学的に適切なデータの処理が可能になります。API は、どのロケールでも同じ（ユニバーサル）ですが、API の動作は、ロケールの設定によって異なります。このような NLSRTL のアーキテクチャは、ソフトウェアのグローバリゼーションの要件を満たすうえで、大きな柔軟性を提供します。

SQL、PL/SQL などによって構築された Oracle データベース・アプリケーションは、NLSRTL および Oracle NLS アーキテクチャをベースに構築され、グローバリゼーション・モデルを継承します。たとえば、PL/SQL で記述されたアプリケーションは、1 つの国際バージョンであらゆるロケールをサポートできます。ロケールのデータは、NLS Runtime Library 機能やその使用方法を変更せずに更新できます。したがって、追加された新規言語のサポート、またはロケールに依存する機能の操作は、データ・コンポーネントを更新することによりカスタマイズできます。エンド・ユーザーやデータベース管理者は、パラメータの設定を変更して、RDBMS の動作を変更しロケールのニーズに適合させることができます。これにより柔軟なアーキテクチャが提供されるので、アプリケーションのコードを変更せずに、言語に依存するデータを変更し新規言語のデータを追加できます。また、単一の製品で複数の言語がサポートされます。

グローバリゼーションとは

グローバリゼーションは、ユーザーの母国語およびロケール・プリファレンスでコンテンツを変更せずにレンダリングして、世界のどこからでも同時にアクセスと実行が可能な多言語のアプリケーションとソフトウェア製品を開発するプロセスです。National Language Support により、顧客はグローバルなアプリケーションとソフトウェアを構築できます。グローバリゼーションの実装には複数の手順が必要です。グローバリゼーションのプロセスは、国際化およびローカライゼーションのプロセスであると言えます。

まず、国際化プロセスについて説明します。アプリケーションをどこでも実行可能にするには、米国式以外のキーボードや国によって異なるハードウェアに対応する、あらゆる言語のオペレーティング・システムで使用可能にする必要があります。文字列に対するアプリケーションの依存性がハードコードされていないこと、US バージョン以外の他の製品と相互運用できることが必要です。翻訳を容易にするには、アプリケーションにより、言語テキストを独立した翻訳可能なファイルに分割する必要があります。アプリケーションがコードセットに依存しないこと、分散環境でマルチバイト・キャラクタを処理し相違点に対処できることが必要です。正しい情報の表示には、ユーザーが指定するロケールを検出する機能が必要です。

ローカライゼーションには、独立したファイル・テキストの翻訳が含まれます。情報は、データの形式、照合、通貨、日時、テキストの方向性を含め、ユーザーの文化的背景に一致した方法でユーザーに表示される必要があります。

Oracle Corporation 発行「Oracle Database 10*g*: The Power of Globalization Technology」の翻訳版です。

グローバリゼーションを必要とする人々

あなたのビジネスは、国際的にアピールする力を持つインターネット上で、製品やサービスを提供していますか。多数の B2B および B2C 企業が Web で販売を行っていますが、大きな潜在市場を見逃してはいませんか。特定の製品またはサービスを探している消費者を考えてみます。この消費者は検索キーワードを母国語で Web ブラウザに入力しますが、該当する企業はほとんどありません。消費者が Web サイトを見つけても、サイトが母国語で表示されていなければ、購入する可能性は統計的にほとんどありません。販売を逸すること自体は想定の範囲内かもしれませんが、設計が不適切なグローバル・アプリケーションではコストが非常に高くなります。国際的な顧客の要求を満たし多数のサーバーとデータベースを維持し、国際的なサイトごとに 1 つのサーバーとデータベースを持つホスト・サービスを検討します。この場合、複雑なデータ・レプリケーションおよび統合が必要になります。各アプリケーションについて、重複するコードの維持が必要です。これでは、メンテナンスの負担が増すばかりです。

なぜグローバリゼーションか

成長する最大の市場セグメントを狙うには、グローバリゼーションが必要です。現在、世界の人口の 92%が英語以外の言語を話すと言われています。当初は 8%の英語を話す人々が、最も初期の熱心な Web ユーザでしたが、この状況は急速に変化しています。Forrester Research 社は、現在、インターネット・ユーザーの 50%のみが英語を話し、2005 年までには、英語を使用するインターネット・ユーザーはわずか 3 分の 1 になると推定しています。英語以外の言語を母国語とする人々が、最も急成長するインターネット・ユーザーのグループであり、消費者ということになります。これについては、次に説明します。最近の調査によると、アメリカの主要な Web サイトでは、外国からの注文の約半分を拒否したことがわかりました。その理由は、連絡先、住所などの多言語データの保存、アクセス、取得が不可能であること、通貨の問題が海外の顧客の意欲を阻害するからです。では、英語以外の言語を話すユーザーは、多言語の情報を入力できる Web サイトから購入するでしょうか。そのサイトにユーザーの母国語およびロケール・プリファレンスでコンテンツが表示されていない場合、購入することは難しいと考えられます。実際、コンテンツが母国語で表示されている場合、消費者が購入する可能性は 4 倍になると言われています。

Oracle Globalization Support の新機能

拡張された Unicode の使用可能性

Unicode は、汎用のエンコード・キャラクタです。これにより、単一のキャラクタ・セットを使用して、あらゆる言語の情報を保存できます。Unicode はプラットフォーム、プログラム、言語に関係なく、すべての文字に対して固有なコード値を提供します。Unicode 標準は、多数のソフトウェアおよびハードウェア・ベンダーによって採用され、現在では、多くのオペレーティング・システム、Web ブラウザが Unicode をサポートしています。XML、Java、JavaScript、LDAP、CORBA 3.0、WML などの新しい標準は、Unicode を必要とします。また、Unicode は、ISO/IEC 10646 標準に準拠しています。

オラクル社は、Oracle 7 で Unicode をデータベース・キャラクタ・セットとしてサポートするようになりました。Oracle 9*i* 以降では、Unicode のサポートが大幅に拡大され、顧客はグローバリゼーションのニーズに対応する適切なソリューションを見つけることができるようになりました。Oracle Database 10*g* は、Unicode 標準の最新バージョンである Unicode 4.0 をサポートします。

Unicode エンコーディング

Unicode 4.0 の文字をエンコードする一般的な方法は 2 つあります。

- UTF-8 エンコーディング
- UTF-16 エンコーディング

UTF-8 エンコーディング

これは、Unicode の 8 ビット・エンコーディングです。0x00 から 0x7F までの文字コードが ASCII と同じ意味を持つ、可変幅のマルチバイト・エンコーディングです。このエンコーディングにおける 1 つの Unicode 文字は、1 バイト、2 バイト、3 バイトのいずれかになります。一般的に、ヨーロッパのスクリプトの文字は 1 バイトまたは 2 バイトで表現される一方、アジアのスクリプトの文字は 3 バイトで表現されます。追加文字は 4 バイトで表現されます。

UTF-16 エンコーディング

これは、Unicode の 8 ビット・エンコーディングで、0x0000 から 0x007F までの文字コードが ASCII と同じ意味を持ちます。このエンコーディングでは、1 つの Unicode 文字は、2 バイトまたは 4 バイトです。ヨーロッパのスクリプトの文字およびアジアのほとんどのスクリプトの文字は、2 バイトで表現されます。追加文字は 4 バイトで表現されます。

Unicode データベース

Oracle データベースは、データベース・キャラクタ・セットという概念を持っています。これが、SQL CHAR データ型で使用されるエンコーディング、および表名、列名、SQL 文などのメタデータを指定します。UTF-8 では、Unicode データベースをデータベース・キャラクタ・セットとして定義する必要があります。UTF-8 エンコーディングを実装する Oracle キャラクタ・セットは 3 つあります。最初の 2 つは ASCII ベースのプラットフォームに対して設計されていますが、3 つ目のキャラクタ・セットは EBCDIC プラットフォームで使用します。

- AL32UTF8
  AL32UTF キャラクタ・セットは、Unicode 4.0 に準拠し、文字を 1 バイトから 3 バイトにエンコードします。追加文字には 4 バイトが必要です。AL32UTF8 は、継続して最新の Unicode 標準に準拠しているため、Oracle Database 10*g* 以降のデータベースで UTF-8 のサポートが必要な顧客に推奨されるキャラクタ・セットです。
- UTF8
  UTF8 キャラクタ・セットは、Unicode 3.0 および CESU-8 に準拠し、文字を 1 バイトから 3 バイトにエンコードします。
- UTFE
  UTFE キャラクタ・セットは Unicode 3.0 に準拠します。EBCDIC プラットフォームでデータベース・キャラクタ・セットとして使用され、UTF-8 エンコーディングをサポートします。

各国語キャラクタ・セット

CREATE DATABASE 文を使用してデータベースを作成する場合、SQL NCHAR データ型に各国語キャラクタ・セットを指定できます。たとえば、WE8ISO8859P1 をデータベース・キャラクタ・セット、AL16UTF16 を各国語キャラクタ・セットとして作成します。これは、SQL NCHAR データ型を使用することで、Unicode データを Unicode 以外のデータベースに保存できることを意味します。

新規 Unicode データ型

SQL NCHAR データ型は Unicode です。データベース全体を変換する必要のない顧客に、多言語サポートの別の手段を提供します。Unicode 文字は、データベース・キャラクタ・セットの設定とは関係なくデータ型の列に保存できます。NCHAR データ型は Oracle9*i* で再定義され、それ以降は Unicode データ型になっています。つまり、Unicode エンコーディングのみでデータを保存します。SQL CHAR データ型の場合と同様に、SQL NCHAR データ型を使用できます。

SQL NCHAR データ型で使用するエンコーディングは、データベースの各国語キャラクタ・セットとして指定されます。各国語キャラクタ・セットとして、次の 2 つの Oracle キャラクタ・セットから指定できます。

- AL16UTF16
  これは、SQL NCHAR データ型のデフォルトのキャラクタ・セットです。このキャラクタ・セットは、Unicode データを UTF-16 エンコーディングでエンコードします。データは 16 ビットの UTF-16 ユニットで数えられます。
- UTF8
  UTF-8 を SQL NCHAR データ型に指定した場合、SQL データ型に保存されたデータは、UTF-8 でエンコードされますが、あたかも AL16UTF16 が指定されているかのように数えられます。デフォルトでは、データは UTF-16 エンコーディングで SQL NCHAR データ型に保存されます。NCHAR 列と NVARCHAR2 列で指定された長さは、常に、バイト数ではなく文字数で表されます。

追加文字（Supplementary Characters）

Unicode は拡張可能で、100 万以上の文字をエンコードできます。拡張された文字は、追加文字と呼ばれます。追加文字は、将来の Unicode 標準の拡張で文字の表現を可能にする設計がされています。UTF-8 および UTF-16 での追加文字は 4 バイト必要です。また、Unicode では使用されない専用の未使用領域があります。この領域により、特殊化したスクリプトの保存が可能です。たとえば、スター・トレックの Web サイトがクリンゴン語をサポートするとします。追加文字を使用して、専用の領域にクリンゴン語の各文字を特別にマッピングできます。

キャラクタ・セマンティクス

キャラクタ・セマンティクスにより、マルチバイト文字列の記憶要件およびアプリケーションの文字列操作を処理するタスクが簡素化されます。Unicode データベースを設定する場合のキャラクタ・セマンティクスを考えてみます。たとえば、ASCII キャラクタ・セットとは異なり、UTF-8 および UTF-16 文字は、マルチバイトのセグメントに保持されます。キャラクタ・セマンティクスを使用する UTF-8 など、可変幅のキャラクタ・セットの場合、文字列の処理は特に容易になります。たとえば、CHAR(10)は、10 コード・ポイントまたは 10 文字の記憶域を意味し、SUBSTR、LENGTH、INSTR などの SQL 文字列関数によってサポートされます。

ロケール適用範囲の拡大

世界の 57 の言語、88 の国と地域、および 200 種類を越えるキャラクタ・セット。Unicode データベースおよびデータ型の使用により、Oracle は、ほとんどの現代言語および文字をサポートされます。Oracle は、特定の地域固有の多様な文化的背景をサポートします。デフォルトの時刻のフォーマット、日付のフォーマット、数字および通貨の表記規則が、各地域設定に基づいて処理されます。異なる NLS パラメータを設定することにより、データベース・セッションは異なる文化の設定を利用できます。たとえば、二重通貨がサポートされます。ドイツでは、ドイツマルクとユーロとの表示が重要な場合があります。

また、日付および時刻のフォーマットは、グローバルなアプリケーション構築での重要な考慮事項です。時刻、日付、月、年に関する世界の様々な表記規則が、地域のフォーマットで処理されます。たとえば、英国では、日付は DD-MON-YYYY フォーマットを使って表示されますが、日本では一般的に YYYY-MON-DD が使用されます。

通貨、貸方、借方の記号など、貨幣や数字も地域のフォーマットで表記することが必要です。基数の記号および 3 桁区切りは、ロケールにより定義されます。たとえば、米国では、小数点の記号はドット「.」ですが、フランスではカンマ「,」です。したがって「$1,234」は、国によって意味が異なります。

世界中で多数の暦法が使用されています。Oracle は次の 7 つの暦法をサポートします：グレゴリオ暦、和暦、台湾暦（中華民国）、タイ仏教暦、ペルシャ暦、英語版イスラム暦、アラビア語版イスラム暦

Oracle のソート機能の概要

Oracle は、様々な言語と文化の複雑なソート要件に対応するソート機能を提供します。言語が異なれば、ソート順も異なります。さらに、同じアルファベットでも文化や国が異なる場合、単語を異なる方法でソートする場合があります。たとえば、デンマーク語では、文字「Æ」は「Z」の後に来ますが、「Y」と「Ü」は、同一の文字の変形と見なされます。ソート順には大文字／小文字を区別するもの、しないもの、アクセントを無視するもの、しないものがあります。東アジアの表意文字の画数や部首による順序付けのように、文字の発音に基づく場合も外観に基づく場合もあります。もう 1 つの一般的なソートの問題は、文字の結合です。たとえば、伝統的なスペイン語では、「ch」は 1 つの独立した文字です。ソート順は、cerveza、Colorado、cheremoya、のようになります。つまり、文字「c」は、後に続く文字が「h」であるかの確認までソートされません。複数の異なる種類のソートを提供する Oracle は、多数の言語を同時処理するよう設計された新しい多言語の ISO 規格（10646）への準拠に加え、言語学的に正しいソートが実行できます。

バイナリ・ソートの使用

通常、文字データが保存される場合、ソート基準は、文字コード体系により定義される文字の数値に基づきます。これは、バイナリ・ソートと呼ばれます。バイナリ・ソートは最速のソートで、ASCII 規格および EBCDIC 規格で A から Z までの文字が数値の昇順に定義されるため、英語のアルファベットに対して妥当な結果が出ます。ただし、ASCII 規格では、すべての大文字が小文字よりも前に表示

されます。EBCDIC 規格では、その逆になり、すべての小文字は大文字よりも前に表示されます。他の言語で使用される文字が存在する場合、バイナリ・ソートでは通常、よい結果が得られません。たとえば、文字コード体系で Ä が B よりも高い数値を持つ場合、昇順の問い合わせ ORDER BY は、文字列 ABC、ABZ、BCD、ÄBC を順に返します。漢字を使用した言語の場合、バイナリ・ソートは言語学的に意味がありません。

言語ソートの使用

アルファベットの文字順に一致するソート基準の作成には、文字コード体系の中で文字の数値から独立して文字をソートする、別のソート技法が必要になります。この技法は、言語ソートと呼ばれます。言語ソートは、数値で文字を置き換えることで機能します。この数値は、各文字の適切な語学上の順序を反映します。これらの数値は、メジャー値とマイナー値を含む表にあります。Oracle は、2 つの段階で文字列を比較します。最初の段階では、メジャー表の文字列全体のメジャー値を比較し、第 2 段階ではマイナー表のマイナー値を比較します。メジャー表の各エントリには、Unicode コードポイントおよびメジャー値が含まれます。通常、同じ外観を持つ文字は、同一のメジャー値を持ちます。Oracle は、メジャー値が同一でマイナー値が異なる、発音区別記号とケースが異なる文字を定義します。Oracle は 2 種類の言語ソートを提供します。ヨーロッパ言語に使用する単一言語ソート、アジアの言語に使用する多言語ソートです。

単一言語ソートの使用

Oracle は、ほとんどすべてのヨーロッパ言語に対して文化固有のソート順を含む単一言語ソートを提供します。多言語ソートの使用により、Oracle は、単一言語ソートを拡張し、1 つのソートの一部として他の言語をソートできます。これは、複雑なソート規則やグローバルな多言語データベースを持つ地域や言語で役立ちます。さらに、Oracle は以前のリリースが定義するすべてのソート順をサポートします。

たとえば、フランス語ソートはサポートされていますが、ソート順をフランス語から French_M に変更することで、多言語のフランス語ソートも適用できます。これにより、ソート順は GENERIC_M ソート順に基づき、第 2 レベルを右から左にソートできます。オラクル社では、表に多言語のデータが含まれている場合、多言語ソートの使用をお薦めします。表にフランス語のみが含まれる場合は、メモリー使用率に注意してフランス語ソートを使用すると、よりよいパフォーマンスが得られます。拡張性とパフォーマンス間にはトレードオフが成立します。

アジアの言語データまたは多言語のデータに対して、Oracle は ISO 規格（ISO14651）および Unicode 規格 4.0 に基づくソート機能を提供します。アジア言語に対する多言語ソートは、画数、ぴん音、部首の数に基づく 3 段階の方式で実装されます。さらに、標準的な等価とサロゲート・コードポイントのペアは、1 回のソートで最大 110 万のコードポイントを定義できる処理容量で実装されます。

言語索引の使用

言語索引の使用により、バイナリ・ソート（最高のパフォーマンスを提供する）とほぼ同じソート・パフォーマンスを達成する一方で、多言語ソートという高度

なソート機能が使用できます。英語以外の言語に対するファンクション索引を作成できます。索引自体は、NLS_SORT が決定する言語ソート順を変更しません。索引は、単にパフォーマンスを向上させます。

多言語索引

多言語の文字データを 1 つのデータベースに保存する場合、ユーザーは 1 つの列に複数の多言語索引を作成する必要があります。この方法は、多言語の特定の列に対する言語ソートのパフォーマンスを向上します。多言語データベースの強力な機能です。

大文字／小文字を区別しない、アクセント記号に依存しない検索

Oracle データベース内の演算は、常に大文字／小文字や、文字のアクセント記号（発音区別記号）のある／なしを区別しますが、大文字／小文字を区別しない、アクセント記号のある／なしを区別しない比較やソートが必要な場合があります。関数 LOWER／UPPER を呼び出して、SQL 文が大文字／小文字を区別しないようにできますが、これはパフォーマンス低下を招きます。Oracle Database 10*g* で導入された大文字／小文字を区別しない、アクセント記号のある／なしを区別しない新しい検索機能は、あらゆる言語ソートに適用でき、パフォーマンスが低下しません。さらに、これは顧客が、既存のコードを変更することなく同じ SQL の動作を使用できる、一挙両得の検索機能です。言語ソートの場合と同様に、大文字／小文字を区別しない検索のパフォーマンスを改善するファンクション索引が構築できます。

正規表現の検索と置換

Oracle の正規表現は、パターンを迅速に記述する簡単で強力なメカニズムを提供し、データベース内のテキストの検索、抽出、書式設定、操作を簡素化します。従来の正規表現エンジンは、英語のテキストのみに対応する設計がされています。ただし、正規表現の実装は、西ヨーロッパのテキストとはまったく異なる文字を持つ様々な言語を網羅します。

Oracle の正規表現の実装は、Unicode 正規表現ガイドラインをベースにしています。REGEXP SQL 関数は、データベース・キャラクタ・セットおよび各国語キャラクタ・セットとしてサポートされるすべてのキャラクタ・セットを処理します。さらに、Oracle は POSIX 正規表現の一致機能を強化し、多言語データを一致させるというユニークな言語学的要件に対処します。

Oracle の正規表現によるパターン一致では、セッション環境から定義される基礎ロケールが考慮されます。これは、大文字／小文字の区別やアクセントによる区別を行うか、文字が一定範囲内にあることを考慮するか、どの照合要素を有効とするかなどの、パターン一致機能のあらゆる側面に影響します。たとえば、ドット（.）が現行のキャラクタ・セットで 1 文字に相当し、データの 1 バイトに相当しないことからもわかるように、エンジンは完全に文字依存です。

日付とタイムゾーン

複数の地域にわたるロケールをサポートするアプリケーションは、タイムゾーンを包括的かつ精密にサポートし、複雑な手動計算を必要としません。日時のデー

タ型は秒以下の正確さで時間のデータを保存できます。日時のデータ型、TSLTZ および TSTZ は、タイムゾーンを認識します。日時の値は、特定のオフセットではなく、特定の地域のローカルタイムとして指定できます。一定の地域のタイムゾーン規則表を使用すると、サマータイムの時間調整を考慮してローカルタイムのタイムゾーン・オフセットが計算され、さらに他の演算に使用されます。

Character Set Scanner ユーティリティ

データ移行の問題

Database Character Set Scanner ユーティリティは、Oracle データベースを新規データベース・キャラクタ・セットに移行する場合の実用可能性と潜在的な問題を評価します。多くの顧客が、Character Set Scanner を使用して、データの切捨て、データベースの Unicode への移行前に拡張を必要とする無効な文字およびフィールドなど、移行の潜在的な問題を見つけ回避しています。新しいデータベース・キャラクタ・セットに移行する場合、Export および Import ユーティリティが元のデータベース・キャラクタ・セットから新規データベース・キャラクタ・セットへのキャラクタ・セット変換を処理できます。ただし、キャラクタ・セット変換は、データ消失またはデータ破損を引き起こす場合があります。たとえば、キャラクタ・セット A からキャラクタ・セット B へ移行する場合、宛先のキャラクタ・セット B はキャラクタ・セット A のスーパーセットであることが必要です。キャラクタ・セット B で使用できない文字は置換文字に変換されます。データ損失を引き置こす可能性があるもう 1 つの使用例は、異なるキャラクタ・セットのデータを含むデータベースをデータベース・キャラクタ・セットのデータから移行することです。どのようにして、このような状況が発生するのでしょうか。クライアントの NLS_LANG キャラクタ・セットの設定がデータベース・キャラクタ・セットと同じ場合、ユーザーは、別のキャラクタ・セットからデータベースにデータを挿入できます。これらの設定が同じ場合、Oracle は送信または受信されるデータが同じキャラクタ・セットであると仮定するため、妥当性チェックと変換が実行されない場合があります。これによって、データに非一貫性の問題が生じる可能性が 2 つあります。1 つの可能性は、データベースが別のキャラクタ・セットからのデータを含んでいるにもかかわらず、両方のキャラクタ・セットに同一のコードポイントがある場合です。2 つ目の可能性は、データベース内に混成のキャラクタ・セットからなるデータがある場合です。たとえば、データのキャラクタ・セットが WE8MSWIN1252 で、ドイツ語と中国語を使用する 2 つの独立した Windows クライアントの両方が、WE8MSWIN1252 に設定された NLS_LANG キャラクタ・セットを使用している場合、データベースには、ドイツ語と簡体字中国語の組合せが含まれます。明らかに、このキャラクタ・セット、つまり WE8MSWIN1252 の場合、中国語の文字は想定されておらずサポートもされません。

データ移行の問題の予測

Scanner がデータベース内のすべての文字データをチェックし、キャラクタ・セットのエンコーディング変更による影響と問題をテストします。スキャン終了時に、データベース・スキャンのサマリー・レポートが生成されます。このレポートは、データベースを新しいキャラクタ・セットに変換するために必要な作業量の見積りを提供します。Scanner は文字データを読み取り、各データ・セルにつき、次の条件をテストします。

- 新しいキャラクタ・セットに変換される場合に、データ・セルの文字コードが変更されますか。
- データ・セルは正常に新規キャラクタ・セットに変換されますか。

- 変換後のデータは現在の列のサイズに適合しますか。

移植可能な NLB データ

Oracle Database 10*g* から、NLB ファイルが移植可能です。Oracle Locale Builder を使用してロケールをカスタマイズ後、生成される NLB ファイルは、たとえば FTP で他のプラットフォームに移植できます。移植された NLB ファイルは、元のプラットフォームで生成された NLB ファイルと同様に使用できます。

Scanner は、CHAR、VARCHAR2、LONG、CLOB、NCHAR、NVARCHAR2、NCLOB 列のみのデータを読み取り、テストします。Scanner は、LONG、CLOB、NCLOB 列に対して、変換後の列サイズのテストを実行しません。

HTML ファイルとプレーン・テキスト・ファイルのスキャン

ファイルを安全にデータベースにロードするには、発生の可能性がある変換についての知識が必要です。Oracle Database 10*g* で導入された Language and Character Set File Scanner（LCSSCAN）ユーティリティにより、ソース・ファイルのエンコーディングを手探りで識別することはありません。LCSSCAN は、統計をベースとするクライアント・ユーティリティで、指定されていないファイル・テキストのキャラクタ・セットと言語を確定します。

Oracle Locale Builder

Oracle Locale Builder により、ユーザーは、わかりやすいグラフィカル・インタフェースを使用して、ほとんどすべてのタイプのロケール定義を単純かつ安全にカスタマイズできます。Oracle Locale Builder は、NLS ロケール・データ定義へのアクセスと簡易で効率的な定義方法を提供します。グラフィカル・ユーザー・インタフェースにより、ユーザーは多様なロケール特有のデータを容易に表示、変更および定義できます。データが、定義ファイル（*.nlt と*.nlb）から抽出され、読込み可能なフォーマットで表示されるので、これらのファイルで使用される定義の独自のフォーマットに煩わされることなく情報を処理できます。結果の出力ファイル形式は、テキスト形式（.nlt）またはデフォルトのバイナリ・フォーマット（.nlb）のいずれかを選択できます。

Oracle Locale Builder は、ローカル・アプリケーションとして実行することにより、ローカルで、または標準 Web ブラウザを使用してリモートでアクセスできます。

Oracle Locale Builder は、次の 4 種類のロケール定義に対応します。

- カレンダの表記規則、日時のフォーマット、数字および通貨の表記規則を含む地域
- 地域の月および曜日の名前、表記方向を含む言語
- キャラクタ・セットの種類、キャラクタ・マッピング、キャラクタ・セット・チャート、および種別を含むキャラクタ・セット
- 言語ソート順、および固有の照合ルールを含む言語ソート

Locale Builder は、新しいロケール定義の追加、または既存の定義のカスタマイズに 2 つの方法を提供します。これを行うには、基本的には継承された設定を構築し、それを変更してプロパティを追加します。たとえば、ハイブリッド言語 French

American を構築します。Locale Builder で、French を選択します。月および曜日の名前、または文字のルールを変更して、一意の French American 言語を作成します。まったく新しい言語の作成とは対照的に、新しい地域をカスタマイズすることが必要な場合があります。特定の言語に対して、新しい地域を定義後、言語の特性に影響を与えずに日時、通貨、数字のフォーマットを変更できます。数字のフォーマットには、小数点、数字を丸める方法、メートル法などの測定の標準が含まれます。通貨記号、小数点、桁区切り記号、貸方、借方の記号などのフォーマットが変更できます。

新しいキャラクタ・セットを作成し、ユーザー定義の文字（UDC）を使用して、既存のエンコードされたキャラクタ・セットの定義を拡張することができます。拡張される文字としては、次のようなものが考えられます。

- 正式名称で使用されている文字（人名漢字など）
- 既存のキャラクタ・セット標準で定義されていない歴史的な漢字
- ベンダー固有の文字
- ユーザーが定義する新しい記号または文字

通常、ユーザー定義の文字は、Unicode および東アジアのキャラクタ・セットの中でサポートされます。ユーザー定義の文字をサポートするキャラクタ・セットでは、少なくとも一組のコードポイントが保留されます。ユーザーは、それらをユーザー定義の文字として使用できます。たとえば、日本語の Shift JIS 文字は、ユーザー定義の文字に対して 1880 コードポイントを保留します。

言語ソートの順は、様々な形に変更できます。コードポイントの範囲は、再構成できます。たとえば、数字を文字の後にソートすること、個々のコードポイントの順を変更することができます。特定の発音区別記号を含む文字すべてに対するソートの変更、または 1 回につき特定の発音区別記号を含む 1 つの文字を変更できます。標準的な等価は、必要に応じて構成またはオフにできます。

Globalization Development Kit

Oracle Globalization Development Kit（GDK）は、一連の Java API から構成されています。GDK を使用することで、アプリケーション開発者は、Oracle による最適なグローバル・インターネット・アプリケーションの開発環境を得ることができます。GDK は、既存の Java グローバリゼーション機能を補完し、中間層の Java アプリケーションと Oracle データベース・サーバー間で、ロケールの動作を同期化します。

GDK が Java に提供する機能は、次の 2 つのカテゴリに分類できます。

- J2EE の GDK アプリケーション・フレームワークは、J2EE ベースのインターネット・アプリケーションを構築するグローバリゼーション・フレームワークを提供します。このフレームワークは、ユーザー・ロケールの判断、ロケール永続性の維持、ロケール情報の処理など、グローバリゼーション・プログラミングに伴う複雑な処理をカプセル化します。一連の Java クラスで構成され、アプリケーションはこれらのクラスを介してフレームワークにアクセスします。これらの関連する Java クラスにより、

アプリケーションはフレームワークと対比してコード化できるため、宣言によりグローバリゼーション動作を拡張できます。

- GDK Java API は、Java での開発をサポートし、Oracle データベース・サーバーと同様、一貫したグローバリゼーション処理を提供します。API は GDK フレームワークに依存しません。GDK フレームワークをベースとしないスタンドアロン型の Java アプリケーションや J2EE アプリケーションからでも、Java API の提供する個々の機能が使用できます。Java API で提供される機能には、日付と数字のフォーマット、Oracle Database と同じ方法を使用するキャラクタ・セットのソートおよび処理が含まれます。また、言語およびキャラクタ・セットの検出テクノロジは、GDK Java API を介して使用できます。

GDK は、JDK バージョン 1.3 以上を実行する Oracle9*i* Database および Oracle Database 10*g* データベースで動作確認され、サポートされています。

中央多言語サーバーの構築

アーキテクチャ

E-Business は、グローバルなコンピュータ・システム上で運営されます。全員が接続されています。そして、すべての情報が一箇所で共有されます。

中央サーバー・アプローチの利用

オラクル社は、1 つの中央サーバーを構築することで、10 億ドル以上を節約することに成功しています。しかし、それよりも以前のオラクル社の状況は、次のようなものでした。情報が何百もの独立したデータベースに分散していることが問題で、マーケティング、セールス、サービスなど、各組織が独自のコンピュータ・システムを持っていました。それぞれのコンピュータ・システムには固有のデータベースがありました。オラクル社は世界中に何百ものデータベースを所有していました。データは断片化され、社員が仕事に必要な情報を見つけるのは困難でした。多数のデータベースが独立していたため、組織間での情報の共有が困難でした。情報を共有できない場合、グループは協力しません。したがって、マーケティングはセールスと協力しませんでした。ドイツは、フランスと協力しませんでした。そして、協力の欠如が、作業の重複と非効率性を招きました。このような非効率性を排除するために、オラクル社は、情報の検索と共有を容易にすることが必要でした。その方法は次のようなものです。E-Business 方式を採用すること。解決策はきわめて単純でした。しかし、E-Business とは何でしょう。すべてはインターネットとグローバリゼーションにあります。E-Business は、グローバル・ネットワークとグローバル・データベースを使用して、取引のあらゆる側面を統合します。ビジネスのすべての機能、つまり、マーケティング、セールス、サプライ・チェーン、顧客サービス、経理、人事管理、すべてが同一のグローバル・ネットワークと同一のグローバル・データベースを使用します。E-Business は、統一されたコンピュータ・システム上で運営されます。全員が接続されています。そして、すべての情報が一箇所にあります。概念的にはシンプルですが、この単一かつ統一されたデータベースへのアプローチには、アプリケーション・ソフトウェアの根本的な変更が必要でした。形態が B2B や B2C でもまたはホス

ティングでも、多くの企業が多言語データに対応し、情報をそれぞれの母国語およびロケールで表示するという課題に直面しています。中間層のソフトウェアは、世界のあらゆる地域からの Web トラフィックをサポートし、顧客の言語とロケール・プリファレンスを識別し、正しく翻訳されたページの作成が必要になります。単一のアプローチにサーバーとデータを統合するなら、すべてのコンポーネントを多言語対応にする必要があります。

データベースでの Unicode ソリューションの実装

中央多言語サーバー構築の第一歩は、あらゆる言語でデータを保存、取得、更新する手段を持つことです。UTF-8 および UTF-16 のサポートにより、世界中のほとんどの現代言語および文字を容易にエンコードできます。これにより、1 つの中央データベース内で多言語を開発、配置、ホスティングできます。多言語の情報を共有でき、なおかつ、各ユーザーが自分のロケール・プリファレンスでサービスの提供を受けることができます。

GDK を備えた単一の多言語アプリケーション・サーバーの実装

多くの場合、アプリケーションは、単一の言語をサポートするように作成されます。システムが拡大すると、多言語サポートに対する要件が発生します。単一の言語をサポートするアプリケーションの多言語バージョンを作成することは、1 つの選択肢ですが、これには多くのリソースとメンテナンス作業が必要です。単一のバイナリおよび単一のコード・ベースから同時に多言語のロケールをサポートすることには、数多くの利点があります。これには、ハードウェア、メンテナンス、および補足的なロケールに対する迅速なサポートのコスト削減が含まれます。Unicode を使用することで、アプリケーションは複数の言語を同時にサポートすることができます。単一の多言語サーバーを使用する方法では、アプリケーション・アーキテクチャのすべての層で Unicode が完全にサポートされます。

顧客は、アプリケーションのグローバリゼーションに関する知識がないため、このソリューションを避けようとします。率直に言って、開発者には、オブジェクト指向のアプリケーションを作成し、企業のビジネス・ロジックを適用するという大きな課題があります。Oracle の Globalization Development Kit（GDK）は、開発プロセスを簡素化し、グローバルな環境をサポートする有効なインターネット・アプリケーションの開発コストを低減します。

Oracle Application Server を使用したアプリケーション管理

Oracle データベースがすべてのデータを管理するのと同様に、Oracle Application Server もすべてのアプリケーションを実行することができます。インターネットおよびイントラネット・アプリケーションに対して、Oracle Application Server は、最も革新的で包括的な中間層サービスを提供します。迅速に Web アプリケーションを開発する独特の機能を提供する一方で、第 2 の主要な特長は、Oracle Application Server と Oracle データベース・テクノロジの密接な統合です。Oracle Application Server の各開発環境が、クライアントのキャラクタ・セットとは関係なく、Unicode データの挿入および取得の複雑性を最小限にするプログラミング方法を提供します。例えば、JSP、Java Servlet、PSP などを検討します。

Oracle Application Server は、Java Servlet の配置を可能にする標準 Java Servlet API を実装します。また、標準 Java ServerPages の仕様に従って、JSP コンパイラを実装します。これにより、ユーザーは標準 JSP を Java Servlet にコンパイルできます。その結果、アプリケーションは、Java（JDK）、Java Servlet、JSP テクノロジで提供されるグローバリゼーション・サポートをフルに活用できます。Oracle JDBC ドライバは、データベース・キャラクタ・セットからのデータを透過的に変換します。すべての Char タイプを、必要であれば UTF-16 および LONG、CLOB に変換します。この透過的な変換の結果、JSP と Oracle JDBC ドライバを呼び出す Java Servlet が、Java 文字列を含むデータベースの列をバインドして定義し、SQL 実行の結果セットの Java 文字列にデータをフェッチします。
PSP および PL/SQL Web ToolKit Oracle iAS は、Web ゲートウェイを提供します。これにより、PL/SQL ストアド・プロシージャが動的な Web コンテンツを生成し、Java Servlet の場合と同様、それをクライアントのブラウザに配信できます。Oracle は、PL/SQL ストアド・プロシージャでインターネット・アプリケーションを開発する、PL/SQL Web Toolkit と呼ばれる API を提供します。API は、Web ページを PL/SQL でフォーマットし、Web ゲートウェイ経由でクライアントのブラウザに送信する手段を提供します。Web Toolkit に加えて、SUBSTRING()、TO_DATE()、LENGTH()など、文字列の操作に、Oracle が提供する SQL 関数を使用できます。VARCHAR、CHAR など、すべての文字列の関数および PL/SQL ストアド・プロシージャ内の SQL 関数は、Unicode で機能します。

PSP は、埋込み PL/SQL コードを持つ HTML ページです。JSP が Java Servlet に関連するのと同様に、PSP は PL/SQL ストアド・プロシージャに関連します。Oracle は、PSP を PL/SQL ストアド・プロシージャにコンパイルして、それらをデータベースにロードする PSP コンパイラを提供します。アプリケーションで PSP および PL/SQL ストアド・プロシージャを使用する場合、データベース内部から SQL または PL/SQL ストアド・プロシージャを使用することにより、自動的に UTF-8 および UTF-16 のデータに直接アクセスできます。

まとめ

柔軟性

Oracle のグローバリゼーション・サポートの使用により、基本的な単一言語、母国語サポートから中央多言語データベースまで、顧客のビジネス・ニーズを満たすロケール環境を作成できます。ローカライゼーションのパラメータ設定は、クライアント・セッションやサーバー、あるいは SQL 関数の中で明示的に指定できます。環境の調整に役立つカスタマイズ機能も使用できます。Unicode のソリューションに移行が必要な顧客は、Unicode データベースまたは Unicode データ型の中からビジネス・ニーズに最適なものを選択できます。

互換性

1 か国で使用するキャラクタ・セット、複数の国で使用するキャラクタ・セット、ベンダー固有のキャラクタ・セットを含む 200 のキャラクタ・セットに対するサポートにより、一貫性と既存のデータとの相互運用性が提供されます。現在、業界では、Unicode 4.0 の完全サポートのために、最も一般的なシングルバイト、マ

ルチバイトおよび固定幅のエンコーディングが使用されています。これには、UTF-8 および UTF-16 が含まれます。Unicode のソリューションに移行が必要な顧客は、Scanner ユーティリティが、互換性の問題を特定し円滑な移行を約束します。オラクル社は、さらに、データベースで自動的かつ透過的に必要なキャラクタ・セットの変換を実行することにより、複数層の多言語アプリケーションの配置をサポートします。Oracle Globalization Development Kit（GDK）には、一連の Java API が含まれています。これらの Java API は、最適なグローバリゼーションのプラクティスおよび Oracle が設計した機能を使用して、Oracle アプリケーションの開発者にグローバル・インターネット・アプリケーションを開発するフレームワークを提供します。GDK は、J2EE の既存のグローバリゼーション機能を補完します。すでに J2EE プラットフォームが、グローバル・アプリケーションを構築する強力な基礎を提供していますが、そのグローバリゼーション機能および動作は、Oracle 製品の機能とは大きく異なります。GDK は、中間層の Java アプリケーションとデータベース・サーバー間の、ロケールにより異なる動作の同期化を可能にします。また、GDK は、最良のデータベース・グローバリゼーション機能を中間層にもたらします。さらに、GDK には、PL/SQL パッケージのスイートが含まれており、PL/SQL で記述されるアプリケーションに対応するグローバリゼーション機能も提供します。

統合

Oracle では、完全にグローバル対応されたシングル・バイナリ・モデルを使用します。これにより、世界共通の Oracle リリースが、確実にユーザーにサービスを提供します。ユーザー・インターフェースは、多数の言語に翻訳されています。多言語アプリケーションと Unicode データベースの構築を行うのであれば、Oracle Application Server は、最も革新的かつ包括的な中間層サービスのセットを提供します。Oracle へのすべてのアクセス・プログラム・インタフェースは、UTF-16 と UTF-8 両方で有効です。したがって、Unicode フォームで記述されたこれらのアプリケーションに優れたネイティブの統合が提供されます。Oracle は、Globalization Development Kit（GDK）も提供します。これは、アプリケーションのグローバル対応の複雑さを排除し、開発者が企業のビジネス・ロジック部分の開発に集中する手助けを行うツールキットです。

Oracle Database 10*g*: **グローバリゼーション・テクノロジの有効性**
2005 **年** 9 **月**
著書: Barry Trute

Oracle Corporation
World Headquarters
500 Oracle Parkway
Redwood Shores, CA 94065
U.S.A.

海外からのお問合せ窓口:
電話: +1.650.506.7000
ファックス: +1.650.506.7200
www.oracle.com