

グローバル化・サポート  
Oracle Unicode データベース・サポート

オラクル・ホワイト・ペーパー  
2005 年5 月

# Oracle Unicode データベース・サポート

概要 .....	3
必要条件 .....	3
Unicodeとは.....	4
補助文字 (Supplementary Characters) .....	4
Unicodeエンコーディング .....	5
UTF-8 エンコーディング .....	5
UCS-2 エンコーディング .....	5
UTF-16 エンコーディング .....	6
UnicodeとOracle .....	7
AL24UTFSS .....	7
UTF8 .....	7
UTFE.....	8
AL32UTF8.....	8
AL16UTF16.....	8
よくある質問と回答 .....	9
結論 .....	13

## 概要

このホワイト・ペーパーでは、ユーザーが様々な Unicode エンコーディングと Oracle のキャラクタ・セットとの関係を理解するために役に立つ情報を提供します。また、Oracle の Unicode 実装に関して最も多く寄せられる質問へも回答します。このホワイト・ペーパーは、ユーザーにとって、グローバリゼーションのニーズを満たす正しい Unicode ソリューションを見つける出発点です。

## 必要条件

同一のアプリケーションまたはデータベースで多言語を扱うことは、長い間、複雑で困難な作業でした。日常の E-Business 運営に必要な条件を処理するために必要な文字を全て含んだ単独のキャラクタ・セットは存在しません。たとえば、EU (European Union: 欧州連合) では、すべての言語をカバーする複数の各国語キャラクタ・セットが必要です。英語のような単一の言語の場合でも、7 ビット ASCII キャラクタ・セットでは、一般に使用されるすべての文字、句読点、技術記号はエンコードできません。

ASCII キャラクタ・セットと各国語キャラクタ・セットには、複数の言語を処理する十分な文字がないという共通の問題があります。各国語キャラクタ・セットの多くは、ASCII をベースにエンコーディングされています。英語とフランス語のような組合せは同じキャラクタ・セットで格納できますが、ドイツ語と中国語、ロシア語と韓国語、さらには繁体字中国語と簡体字中国語など、同じキャラクタ・セットで格納できない組合せがあります。また、異なるキャラクタ・セット間で競合が発生する可能性もあります。なぜなら、同じ数値コードを使用して異なるキャラクタ・セットを表現することもあれば、同じ文字を表現する場合でもキャラクタ・セットによって数値コードが異なることもあるからです。

既存のキャラクタ・セット・エンコーディングの制約を克服するため、1980 年代後半にいくつかの組織がグローバル・キャラクタ・セットの作成に着手しました。このニーズは、1990 年代中ごろのワールド・ワイド・ウェブ (WWW) の開発によってさらに拡大しました。インターネットによるビジネスの手法の変化で、グローバル・マーケットが重視されるようになりユニバーサル・キャラクタ・セットが重要な要件の 1 つになりました。このグローバル・キャラクタ・セットに求められるのは、主要な現行スクリプトのすべてを含んでいること、従来のデータと実装をサポートすることに加え、製品の 1 つの実装ですべての言語に対応できるほど単純なことです。さらに、多言語のユーザーと組織をサポートし、国際標準に準拠し世界各国でデータ交換を可能にすることも要求されます。このようなグローバル・キャラクタ・セットはすでに存在し、広く使用されています。それが Unicode です。

## Unicode とは

Unicode は、ユニバーサルなエンコード・キャラクタ・セットで、あらゆる言語の情報を格納できます。Unicode は、各文字のプロパティを定義し、スクリプトの動作を標準化します。また、双方向テキストの標準アルゴリズムを提供し、他の標準とのクロスマッピングを定義します。Unicode では、プラットフォーム、プログラムまたは言語に関係なくすべての文字に一意のコード値が指定されます。

Unicode 標準は主に言語ではなくスクリプトをエンコードします。ラテン・スクリプトのように、スクリプトの多くは 1 つのスクリプトによって数十種類または数百種類の言語が記述されます。一方、ハングルのように 1 つの言語のみが採用されるスクリプトもあります。

今日、Unicode は、すべての主要言語で記述する際に使用される文字のコードを定義します。Unicode は、ラテン文字、ギリシャ文字、キリル文字、アルメニア文字、ヘブライ文字、アラビア文字、シリア文字、ターナ文字、デーバナーガリー文字、ベンガル文字、グルムキー文字、オリヤー文字、タミール文字、テルグ文字、カンナダ文字、マラーヤラム文字、シンハラ文字、タイ文字、ラオ文字、チベット文字、ミャンマー文字、グルジア文字、ハングル文字、エチオピア文字、チェロキー文字、カナダ先住民の音節、オガム文字、ルーン文字、クメール文字、モンゴル文字、漢字（日本語、中国語、韓国語表意文字）、ひらがな、カタカナ、注音符、イ文字の各スクリプトによって記述されるすべての言語に対応します。

また、句読点記号、発音記号、数学記号、技術記号、音楽記号、矢印、飾り文字なども含まれます。Unicode 標準（バージョン 4.0）は、世界のアルファベット、表意文字、シンボルの中から合計で 96,382 文字を表現します。

多くのソフトウェア・ベンダーやハードウェア・ベンダーが Unicode 標準を採用しています。今では、多くのオペレーティング・システムとブラウザで Unicode がサポートされるようになりました。Unicode は、XML、Java、JavaScript、LDAP、CORBA 3.0、WML などの最先端の標準において必須です。また、Unicode は、ISO/IEC 10646 標準とも同調しています。詳細は、次の Unicode 標準の公式 Web サイト（英語）を参照してください。

<http://www.unicode.org/unicode/standard/standard.html>

## 補助文字（Supplementary Characters）

Unicode の最初のバージョンでは、2 バイトのエンコーディング形式が使用されました。すなわち、いずれのコード・ポイントにも 16 ビットを使用し合計で 65,536 文字を表現できました。現在世界中で使用されるすべての文字を表現するためには、これでも十分ではありません。たとえば、中国語圏では 55,000 文字以上が使用されています。中国語、日本語、韓国語といった言語の場合、エンコードされない表意文字がまだ何万文字もあります。さらに、その多くがほとんど使用されない文字であるにもかかわらず、電子的な保存が必要な文書中に使用されています。

この要件を満たすために、Unicode 標準では補助文字を定義しています。1,024 個ずつの 2 つのグループに分けた 2,048 個のコード・ポイントを割り当て、2 つ一組で補助文字を表現します。各グループから 1 つを取り出し、それを 1 つのペアとして 1 文字を表現します（これをサロゲート・ペアと呼びます）。この組み合わせ

で、合計 1,048,576 文字を定義できます。まず、44,944 個の補助文字が Unicode 3.1 に追加されました。これによって Unicode 標準は複雑になりましたが、多くの異なるコード体系の管理に比べればはるかに簡単です。

## Unicode エンコーディング

多くのテクノロジー同様、Unicode 標準の実装は 1 つのみではありません。一般的な Unicode エンコーディング形式、つまり文字をバイナリ・コードによって表現する方法には、次の形式があります。

- UTF-8
- UCS-2
- UTF-16

異なる Unicode コード体系間の変換は、Unicode 標準で定義される単純なビット単位の演算です。この変換はアルゴリズムをベースにしているため、コストのかかるマッピング・テーブルは必要ありません。

## UTF-8 エンコーディング

UTF-8 は、Unicode の 8 ビット・エンコーディングです。可変幅のエンコーディングであり、7 ビット ASCII の厳密なスーパーセットです。そのため、7 ビット ASCII に含まれるすべての文字が、それぞれまったく同じコード・ポイント値で UTF-8 に存在します。このエンコーディングでは、1 つの Unicode 文字は、1 バイト、2 バイトまたは 3 バイトです。ヨーロッパのスキプトの文字は 1 バイトまたは 2 バイトで表現され、アジアのスキプトの文字は 3 バイトで表現される一方、補助文字は 4 バイトで表現されます。

UTF-8 は、UNIX プラットフォーム、HTML および多くのインターネット・ブラウザで使用される Unicode エンコーディングです。

UTF-8 の主な利点は次のとおりです。

- ヨーロッパのスキプトの場合、必要な記憶域がコンパクトになります。UTF-8 は 7 ビット ASCII の厳密なスーパーセットであるため、一般的にディスクやメモリを占有するデータの記憶域が小さくなります。
- 移行が簡単です。UTF-8 は 7 ビット ASCII データと同じであるため、ASCII ベースのキャラクタ・セットと UTF-8 の変換に必要な労力が大幅に軽減されます。

## UCS-2 エンコーディング

UCS-2 エンコーディングは、Unicode の 16 ビット固定幅エンコーディングで、各文字のサイズはいずれのスキプトの場合も 2 バイトです。UCS-2 は、Java および Microsoft Windows NT 4.0 で使用されている Unicode エンコーディングです。UCS-2 では、Unicode 3.0 までに定義された Unicode 文字のみがサポートされ、補助文字のサポートはありません。

UCS-2 の主な利点は次のとおりです。

- アジア系言語のスキプトの場合、すべての文字が2バイトで表現されるため、必要な記憶域がコンパクトになります。
- すべての文字の幅が同じため、文字列処理が高速になります。
- Java クライアントや Microsoft クライアントとの互換性が高まります。

### UTF-16 エンコーディング

UTF-16 エンコーディングは、Unicode の 16 ビット・エンコーディングです。UTF-16 は基本的に UCS-2 の拡張で、UCS-2 コード・ポイントのペアとして Unicode 3.1 で定義された新しい補助文字のサポートを提供します。

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。ヨーロッパのスキプトの文字 (ASCII を含む) およびアジアのほとんどのスキプトの文字は、2 バイトで表現されます。補助文字は 4 バイトで表現されます。UTF-16 は、Microsoft Windows 2000、XP、2003 で使用される主要な Unicode エンコーディングです。

UTF-16 の利点は次のとおりです。

- アジア系言語のスキプトの場合、必要な記憶域がコンパクトになります。UTF-16 の場合、アジア系言語で一般的に使用されている文字の大部分が 2 バイトで表現されるため (UTF-8 の場合、同じ文字に対し 3 バイト必要)、ディスクやメモリで占有される記憶域が小さくなります。
- Java クライアントや Microsoft クライアントとの互換性が高まります。

図 1 に、UTF-8、UCS-2、UTF-16 の各エンコーディングで定義されている文字の一部とその文字コードを示します。最後の文字は、Unicode 3.1 に追加された補助文字の 1 つ、ト音記号 (音楽符号) です。

文字	UTF-8	UCS-2	UTF16
A	41	0041	0041
c	63	0063	0063
Æ	C3 86	00C6	00C6
Ö	C3 B6	00F6	00F6
گ	DA B2	06B2	06B2
亜	E4 BA 9C	4E9C	4E9C
♫	F0 9D 84 9E	該当なし	D834 DD1E

## Unicode と Oracle

オラクル社は、Oracle 7 から Unicode をデータベース・キャラクタ・セットとしてサポートしています。次の表に、Oracle でサポートされている Unicode キャラクタ・セットをまとめます。

キャラクタ・セット	RDBMS でのサポート	Unicode エンコーディング	Unicode のバージョン	データベース・キャラクタ・セット	各国語キャラクタ・セット
AL24UTF8	7.2 - 8i	UTF-8	1.1	対応	非対応
UTF8	8.0 - 10g	UTF-8	2.1 <sup>1</sup> 3.0 <sup>2</sup>	対応	対応*
UTFE	8.0 - 10g	UTF-8	2.1 <sup>1</sup> 3.0 <sup>2</sup>	対応	非対応
AL32UTF8	9i - 10g	UTF-8	3.0 (9iR1) 3.1 (9iR2) 3.2 (10gR1) 4.0 (10gR2)	対応	非対応
AL16UTF16	9i - 10g	UTF-16	3.0 (9iR1) 3.1 (9iR2) 3.2 (10gR1) 4.0 (10gR2)	非対応	対応*

<sup>1</sup> RDBMS 8.0.x から 8.1.6 まで

<sup>2</sup> RDBMS 8.1.7 から 9iR2 まで

\* Oracle 9i および Oracle Database 10g のみ

### AL24UTF8

AL24UTF8 は、Oracle によってサポートされた最初の Unicode キャラクタ・セットです。RDBMS 7.2 で Unicode キャラクタ・セットとして導入されました。

AL24UTF8 は、マルチバイト Unicode キャラクタ・コード体系 UTF-8 の頭字語です。AL24UTF8 コード体系は、現在では廃止された Unicode 1.1 に基づいています。AL24UTF8 は Oracle 9i 以降サポートされていません。既存の AL24UTF8 データベースの移行パスは、Oracle 9i にアップグレードする前の UTF8 へのアップグレードです。

### UTF8

UTF8 は、Oracle 8 および 8i の UTF-8 エンコード・キャラクタ・セットでした。これが Oracle の標準ネーミング規則 (<Language><bit size><encoding>) に従わない最初のキャラクタ・セットです。ネーミング規則に従うと、このキャラクタ・セットの名前は AL24UTF8 (AL は All Languages の略) となります。

当時、このキャラクタ・セットが Oracle の最後の UTF-8 エンコード・キャラクタ・セットになると考えられていました。Oracle 8.0 から 8.1.6 までは Unicode 2.1 に従い、8.1.7 と 9i では Unicode 3.0 にアップグレードされました。既存のインストールとの互換性を維持するため、Oracle の今後のリリースでもこのキャラクタ・セットは Unicode 3.0 のままです。

特定の補助文字は、Unicode 3.1 まで Unicode に割り当ててはありませんが、補助文字の割り当ては、すでにバージョン 3.0 で定義されています。そのため、補助文字が UTF8 データベースに挿入されても、データベース内の実際のデータは壊れません。6 バイトの記憶域を占有する 2 つの分離した未定義文字として扱われます。オラクルでは、補助文字を完全にサポートするために、AL32UTF8 または AL16UTF16 に切り替えることをお勧めします。

## UTFE

UTFE は、EBCDIC プラットフォームの UTF8 データベース・キャラクタ・セットです。ASCII ベースのプラットフォームの UTF8 と同じプロパティを持ちます。

EBCDIC Unicode変換形式については、Unicode Technical Report #16 - UTF-EBCDIC で説明しています。最新のUnicode Technical Reportsの一覧は、<http://www.unicode.org/unicode/reports/> (英語) を参照してください。

## AL32UTF8

AL32UTF8 は、Oracle9i で導入された UTF-8 エンコード・キャラクタ・セットです。AL32UTF8 は、Unicode 標準の最新バージョンをサポートするデータベース・キャラクタ・セットで、新しく定義された補助文字もサポートします。補助文字はすべて 4 バイトで格納されます。

## AL16UTF16

AL16UTF16 は、Oracle の最初の UTF-16 エンコード・キャラクタ・セットです。デフォルトの各国語キャラクタ・セットとして Oracle9i Database で導入され、Oracle Database 10g まで採用されました。各国語キャラクタ・セットが SQL NCHAR データ型 (NCHAR、NVARCHAR2 および NCLOB) のキャラクタ・セットを決定するため、データベース・キャラクタ・セットによっては SQL CHAR データ型 (CHAR、VARCHAR2、LONG および CLOB) のエンコーディングに影響します。

AL16UTF16 は、Unicode 標準の最新バージョン (4.0) をサポートする各国語キャラクタ・セットで、新しく定義された補助文字もサポートします。補助文字はすべて 4 バイトで格納されます。



## よくある質問と回答

### Unicode にマイナス面はありますか？

Unicode には多くの利点がありますが、その使用にはいくつかの犠牲もあります。たとえば、データをシングル・バイト・キャラクタ・セットで表現できる場合、UTF-16 エンコーディング形式を使用すると 2 倍の記憶域が必要になります。現在、アジア系言語のダブルバイト文字を格納しているデータベースを UTF-8 に移行する場合、1 文字あたりに必要な記憶域が 2 バイトから 3 バイトへと 50% 増加します。いずれの場合も、データの拡張により、ネットワークでやりとりするデータの量やメモリ内に保持されるデータの量が増加します。データの拡張の割合は、使用する言語の要件に最適な Unicode エンコーディングを選択することで大幅に抑えることができます。

また、Unicode などのマルチバイト・キャラクタ・セットは、文字列操作のパフォーマンスが多少低下します。これは、マルチバイト・エンコーディングの特質のためです。1 文字の表現に必要なバイト数が増えるため、文字処理アルゴリズムでは先読みして文字の境界場所を判断する必要があります。

オラクル社は、すべてのデータベースのリリースで Unicode の機能向上に取り組んでいます。通常、Unicode データベースは同じリリースのシングルバイト・データベースとほとんど同じ機能を提供します。SQL 文字列処理機能が集中的に使用される環境では、シングルバイト・データベースよりも Unicode データベースの方がパフォーマンスは大きく影響を受けます。すでにマルチバイト・データベース・キャラクタ・セットを使用している場合は、Unicode に移行しても処理のパフォーマンスが同レベルまたはそれ以上になります。

### Unicode はどういう場合に使用すればよいですか？

オラクル社では、あらゆる新規システムの展開で Unicode を使用することをお勧めしています。また、既存システムを Unicode に移行することもお勧めします。システムの Unicode での展開には、有用性、互換性、拡張性の面で様々な利点があります。

現在、多言語のデータをサポートする必要がない場合または Unicode が必要ない場合でも、長期的に見れば、Unicode は新規システムに最適な選択肢で、最終的には時間と費用の節約になると同時に競争力の面でメリットをもたらします。

### Unicode を採用する場合は、Unicode データ型または Unicode データベースを使用する必要がありますか？

UTF-8 でエンコードされた文字を SQL CHAR データ型 (CHAR、VARCHAR2、CLOB および LONG) として格納できる Unicode データベースを作成できます。Unicode のサポートを徐々に拡大したい場合や、多言語データを特定な列でのみサポートする必要がある場合、Unicode データを SQL NCHAR データ型 (NCHAR、NVARCHAR2 および NCLOB) で格納できます。SQL NCHAR データ型は Unicode データの格納にのみ使用されるため、Unicode データ型と呼ばれます。

ビジネスのニーズに最適な Unicode ソリューションを決定する際に、多くの要素を考慮する必要があります。状況によっては、1つのデータベースに Unicode データベースと Unicode データ型の両方をデプロイすると有効です。このトピックの詳細は、『Oracle Database 10g グローバリゼーション・サポート・ガイド』の「Unicode を使用した多言語データベースのサポート」を参照してください。

**Oracle9i 以前のデータベースで NCHAR を使用して Unicode を格納できますか？**

できません。Oracle8 と Oracle8i の NCHAR 型は、Oracle 独自の特別な固定幅のアジア・キャラクタ・セットをサポートするためのものです。アジア・キャラクタ・セットの処理パフォーマンスを向上させるために採用されました。これらのキャラクタ・セットの例として、JA16EUCFIXED、JA16SJISFIXED、ZHT32EUCFIXED などがあります。

Oracle9i 以前のリリースでは、各国語キャラクタ・セットとしてサポートされている Unicode キャラクタ・セットはありません。

**Unicode データ型をサポートするために、どの Unicode エンコーディングを選択すればよいですか？**

このトピックの詳細は、『Oracle Database 10g グローバリゼーション・サポート・ガイド』の「Unicode を使用した多言語データベースのサポート」を参照してください。

**既存の列を Unicode データ型に移行する最適な方法を教えてください。**

このトピックについては、次のOTNサイトにあるホワイト・ペーパー「**多言語データベース/アプリケーションを目的としたUnicodeデータ型 への移行**」を参照してください。

<http://otn.oracle.co.jp/tech/globalization/index.html>

**UCS-2 と UTF-16 の違いは何ですか？**

UCS-2 は、各 Unicode 文字が 2 バイトで表現された固定幅の Unicode エンコーディングです。UTF-16 は、UCS-2 の厳密なスーパーセットで、最新の Unicode 標準で定義される補助文字 (1 文字あたり 4 バイトが必要なサロゲート・ペア) をサポートします。

**データベース・キャラクタ・セットとして AL16UTF16 を使用できないのはなぜですか？**

データベース・キャラクタ・セットは、SQL、SQL メタデータおよび PL/SQL ソース・コードの識別と保持に使用されます。EBCDIC またはシングル・バイトの 7 ビット ASCII のうち、配置プラットフォームにサブセットとしてネイティブなコードを持つ必要があります。そのため、固定幅のマルチバイト・キャラクタ・セット (UTF-16 など) をデータベース・キャラクタ・セットとして使用することはできません。

## UTF-8、UTF8、AL32UTF8の違いは何ですか？

UTF-8は可変幅のエンコーディングであり、7ビットASCIIの厳密なスーパーセットです。UTF-8では、1つのUnicode文字は1バイト、2バイト、3バイトまたは4バイトです。UTF8は、Unicode 3.0までのUTF-8エンコーディングのみをサポートします。AL32UTF8は、最新のUnicode標準で定義された補助文字をサポートするOracle Unicode キャラクタ・セットです。

Oracle社はOracle9iで補助文字をサポートするUTF8の定義をアップグレードせずに、他のUTF-8 キャラクタ・セットを作成していますが、それはなぜですか？

UTF8は最初、Oracle 8.0にさかのぼって定義されました。補助文字の概念がないため、1文字あたり最大で3バイトでした。

UTF8の定義をOracle9iで変更した場合、Oracle9i以前の古いUTF8定義で実行するクライアントで、下位互換性の問題が発生します。

## Oracleの次のメジャー・リリースでUnicode データベース・キャラクタ・セットを再度変更する必要がありますか？

それは、Unicode 標準の今後の方向によって大きく異なります。サロゲート・ペアの導入により、100万文字以上を補助文字として定義できます。この規模は世界のすべての言語を十分カバーできるため、Unicode 標準の将来のバージョンでアーキテクチャの大幅な変更はないと思われます。

Oracle社では、Unicodeの今後のバージョンをサポートするために必要に応じて、AL32UTF8とAL16UTF16を拡張する予定です。

## 補助文字は日常的によく使用されますか？

45,000文字を超える補助文字の大部分は、CJK（中国語、日本語、韓国語）文字の中でほとんど使用されない文字です。残りの文字は、ゴート、古いイタリック、音楽符号、数学記号です。サポートの必要な言語とアプリケーションの性質によっては、これらの文字が日常的に使用されません。しかし、これらの文字がオペレーティング・システムで普通に使用できる場合、ユーザーがこれらの文字を挿入するのを防ぐことはできません。

## UTF8 データベースに補助文字が挿入されるとどうなりますか？

6バイトの記憶域を占有する2つの分離した未定義文字として扱われます。補助文字はUTF8データベースで格納と検索ができますが、これらの文字を操作すると予期しない結果になる場合があります。また、6バイトのUTF8文字は正式なUTF-8エンコーディングとして認識されないため、Java、HTMLなど、他のテクノロジ・スタックとの通信で問題が発生することがあります（この問題は、クライアントのNLS\_LANG キャラクタ・セットをAL32UTF8に設定し、6バイト形式を4バイト形式に変換させることで回避できます）。

この補助文字変換形式については、Unicode Technical Report #26 - Compatibility Encoding Scheme for UTF-16: 8-Bit (CESU-8)で説明されています。最新のUnicode Technical Reportsの一覧は、<http://www.unicode.org/unicode/reports/>（英語）を参照してください。

オラクル社では、補助文字を完全にサポートするために、AL32UTF8 または AL16UTF16 に切り替えることをお勧めします。

#### UTF8 から AL32UTF8 にアップグレードするにはどうしたらよいですか？

AL32UTF8 は UTF8 の厳密なスーパーセットです。そのため UTF8 から AL32UTF8 へのアップグレードは、Oracle9i Database で ALTER DATABASE CHARACTER SET コマンドを使用するか、または CSALTER スクリプトを Oracle Database 10g で実行して行います。ただし、ごくまれに UTF8 データベースに 6 バイトの UTF8 補助文字がある場合、それらのデータ・セルのエクスポートとインポートを実行し、6 バイトの補助文字を 4 バイト形式に変換する必要があります。

一般的なオペレーティング・システムの多くが現時点で補助文字の入力と出力をサポートしていません。そのため、6 バイトで表現された補助文字が現在のデータベースに存在するとは考えられません。ただし補助文字が広くサポートされるようになれば、変わる可能性があります。

UTF8 データベースに存在する 6 バイトの補助文字を調べるには、Character Set Scanner を使用したデータ検証が役立ちます。『Oracle Database グローバリゼーション・サポート・ガイド』の「Character Set Scanner ユーティリティ」を参照してください。

#### Oracle9i で AL24UTFFSS がサポートされなくなったのはなぜですか？

AL24UTFFSS は、現在では廃止された Unicode 1.1 に基づく UTF-8 コード体系をサポートする Unicode キャラクタ・セットとして、Oracle7 で導入されました。

Oracle9i でサポートされている Unicode データベース・キャラクタ・セットは、Unicode 3.0 をサポートする AL32UTF8 と UTF8 です。既存の AL24UTFFSS データベースの移行パスは、Oracle9i にアップグレードする前の UTF8 へのアップグレードです。

#### Oracle8 で、ALTER DATABASE CHARACTER SET コマンドを使用して AL24UTFFSS から UTF8 へ移行できますか？

いいえ、Unicode 2.0 以降では Unicode 1.1 と下位互換性がありません。Unicode 1.1 で定義された韓国ハングル文字のコード値は、その後のバージョンで割り当てが変更されました。Unicode 2.1 で定義されたこの変更をサポートするため、Oracle8 で新しい Unicode キャラクタ・セット (UTF8) が作成されました。

エクスポート・ユーティリティとインポート・ユーティリティは AL24UTFFSS と UTF8 でサポートされた、異なる Unicode 標準間のデータ変換を処理します。

AL24UTFFSS データベースに含まれるハングル文字を調べるには、Character Set Scanner を使用したデータ検証をお勧めします。『Oracle Database グローバリゼーション・サポート・ガイド』の「Character Set Scanner ユーティリティ」を参照してください。

AL24UTFSS から AL32UTF8 への移行についてはどうですか？

AL24UTFSS は Oracle9i またはそれ以降では有効なキャラクタ・セットではありません。そのため Oracle8i での UTF8 への移行パスは、AL32UTF8 にアップグレード前の UTF8 へのアップグレードです。

もう 1 つの方法として、移行前の AL24UTFSS.dmp ファイルを AL32UTF8 データベースにインポートします。

Oracle9i より前のデータベース・クライアント・ソフトウェアを使用して、2 つの新しい Unicode キャラクタ・セット、AL32UTF8 および AL16UTF16 と通信できますか？

これらのキャラクタ・セットを使用して Oracle9i のインスタンスに接続する Oracle8/8i クライアントには、影響を受ける相互運用性の問題があります。これらのキャラクタ・セットは Oracle8/8i には存在しないため、Oracle9i とそれ以降で AL32UTF8 (CHAR 型、VARCHAR2 型、CLOB 型、LONG 型の列) と AL16UTF16 (NCHAR 型、NVARCHAR2 型、NCLOB 型) を使用して格納されたデータを認識するために、アプリケーションにクライアント側のパッチが必要です。

特定のプラットフォームやバージョンのパッチについては、弊社サポート・サービスにお問い合わせください。

## 結論

長い間、Unicode のサポートがオラクル社の顧客に提供されてきました。オラクル社は、Oracle9i Database からその広範囲にわたる Unicode 機能に、最も一般的な 2 種類のエンコーディング形式である UTF-8 と UTF-16 のサポートを追加しました。2 つの新しい Unicode キャラクタ・セットを追加したことで、Unicode 3.1 で導入された補助文字がサポートされました。Oracle Database 10g では、Unicode 4.0 がサポートされています。

Unicode 標準の進化に伴い、オラクル社は一貫して製品を最新のニーズに合わせて拡張しています。顧客のグローバル化のニーズに広く対応するため、データベースの各リリースで新しい Unicode 機能が導入されています。



グローバルゼーション・サポート Oracle Unicode データベース・サポート

2005 年 5 月

著者: Simon Law

寄稿者:

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](http://www.oracle.com)

Copyright © 2005, Oracle. 無断転載を禁ず。

この文書はあくまで参考資料であり、掲載されている情報は予告なしに変更されることがあります。

オラクル社は、本ドキュメントの無謬性を保証しません。また、本ドキュメントは、法律で明示的または暗黙的に記載されているかどうかに関係なく、商品性または特定の目的に対する適合性に関する暗黙の保証や条件を含む一切の保証または条件に制約されません。オラクル社は、本書の内容に関していかなる保証もいたしません。また、本書により、契約上の直接的および間接的義務も発生しません。本書は、事前の書面による承諾を得ることなく、電子的または物理的に、いかなる形式や方法によっても再生または伝送することはできません。

Oracle、JD Edwards、PeopleSoft は、Oracle Corporation および関連会社の登録商標です。他の製品名は、それぞれの所有者の商標です。