

レコードの文字コードの調査

データの文字コード、使用エンコーディング、データ長の調査

```
SELECT  ename , dump(ename , 1016) from  emp  where  empno = 1 ;
```

```
ename      dump(ename , 1016)
```

```
-----
```

```
愛川こずえ
```

```
Typ=1 Len=10 CharacterSet=JA16SJISTILDE: 88,a4,90,ec,82,b1,82,b8,82,a6
```

文字コードについて

Oracle 定義

varchar2 (32) 文字型 32 バイト ← **文字数ではない**ので注意

UTF-8 ASCII 文字 1 文字 : 1 バイト
基本言語文字 1 文字 : 2 バイト
日本語 1 文字 : **3** バイト
補助面文字 1 文字 : 4 バイト

文字セットとエンコーディングの関係

文字セット : コンピュータ上で表せる文字集合を定義したもの (文字の形情報)
出力の表現 (形) だけを示す

例 JIS X 0208、Unicode (日本語表記の文字セットは、この 2 つだけ)
ISO-2022
 ↙ ↘
UCS-2、UCS-4

エンコーディング : 文字をコンピュータ内でどう格納するか**のバイト列の定義**

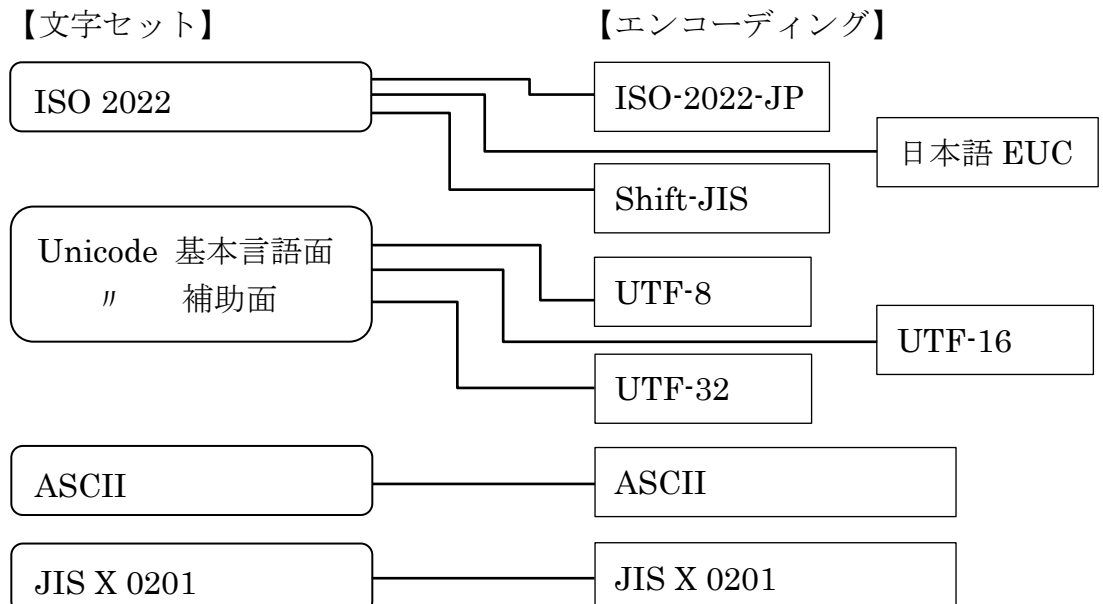
例 Shift-JIS、JIS、EUC-JP、ISO-2022JP、
UTF-8、UTF-16

注) 文字セットとエンコーディングが同時に定義されているものがある

例 ASCII、
JIS X 0201 ローマ字 (ISO 646)、
JIS X 0201 カタカナ (ISO 646)

文字セットとエンコーディングの対応

「エンコーディング」を使用してデータの保存を行い、「文字セット」を用いてデータの内容を表現する



Oracle が使用する文字セットとエンコーディングの組合せ

	文字の形指定	内部格納バイトコード	
Oracle バージョン	Unicode バージョン	設定用エンコード パラメータ	説 明
9.2	3.1	AL32UTF8	UTF-8
		AL32UTF16	UTF-16
		JA16SJIS	Shift-JIS
		JA16SJISTILDE	TILDE 系 Shift-JIS
		JA16EUC	EUC
		JA16EUCTILDE	TILDE 系 EUC
	3.0	UTFE	EBCDIC CESU 準拠 UTF-8
		UTF8	CESU 準拠 UTF-8 補助面非対応
9.0.1	3.0	AL32UTF8	UTF-8
		AL32UTF16	UTF-16
		UTFE	EBCDIC CESU 準拠 UTF-8
		UTF8	CESU 準拠 UTF-8 補助面非対応

Oracle インストール時の設定

文字コード・キャラクタの選択基準

条 件	選択エンコード設定
日本語以外のマルチバイト文字も使う	AL32UTF8 (=UTF8)
日本語と ASCII を使う	AL32UTF8 (=UTF8)
過去のデータを移行する必要がある	
既存データ : Shift-JIS →	JA16SJISTILDE
EUC-JP →	JA16EUCTILDE

Oracle 文字変換の仕組み

Oracle サーバーとクライアントで使用しているエンコードが異なる場合は、
クライアント側の NLS ライブラリにて文字変換が行われる

Unicode 形情報を記録し再現するための、バイト列定義（エンコーディング）の種類

文字セット	エンコーディング
Unicode ↙ ↘ UCS-2、 UCS-4	UTF-32 4 バイト固定長 通常は、使用されない
	UTF-16 2 or 4 バイト アプリとの親和性が低い
	UTF-8 3 バイト アプリとの親和性が高い
	ASCII
	JIS X 0201 ローマ字
	JIS X 0201 カタカナ

JIS X 2028 形情報、もしくは ISO 2022 形情報を記録し再現するための、バイト列定義（エンコーディング）の種類

文字セット	エンコーディング
JIS X 2028 もしくは ISO 2022	Shift-JIS
	日本語 EUC
	ISO 2022 -JP
	ASCII
	JIS X 0201 ローマ字
	JIS X 0201 カタカナ

外字文字への対応

条件		外字 対応状況
DB キャラクター設定 (エンコーディング)	クライアント・アプリ	
Unicode	すべて	対応
SJIS EUC	JAVA 以外	対応
	JAVA	ブラウザにて、 Charset = MS932 (Shift-JS 使用に限る)