

# 文字セットの変換がサポートされていない というエラーを解決したい 技術を知る

- |   |                             |                                |                                 |   |
|---|-----------------------------|--------------------------------|---------------------------------|---|
| <input checked="" type="checkbox"/> 導入／環境設定 | <input type="checkbox"/> 移行 | <input type="checkbox"/> 性能    | <input type="checkbox"/> チューニング | <input type="checkbox"/> バックアップ／リカバリー   |
| <input type="checkbox"/> 冗長化／負荷分散           | <input type="checkbox"/> 監視 | <input type="checkbox"/> データ連携 | <input type="checkbox"/> 災害対策   | <input checked="" type="checkbox"/> 豆知識 |

## 実現方法

PostgreSQL のデータベースサーバーを起動した後や、クライアントからデータベースサーバー上のデータベースへの接続時に、データベース側とクライアント側の文字セット（エンコーディング）の組み合わせによっては、エラーメッセージ「変換はサポートされていません」が出力されることがあります。PostgreSQL にはデータベースサーバーとクライアント間で自動的に文字セットを変換する機構（自動文字セット変換）がありますが、データベース側に指定した文字セットと、クライアント側に指定した文字セットが、サポートされていない組み合わせになっていることが原因です。

このような場合、実際に扱うデータの文字セットに合わせて、正しい組み合わせに設定する必要があります。

## 実行例

クライアント側の文字セットを正しく設定します。なお、ここではクライアントのエンコーディングは SJIS であることを前提とします。

- データベースクラスタ内のデータベースの文字セット（エンコーディング）を確認

```
postgres=# \l
```

### 【出力例】

データベースクラスタ内には、エンコーディングが EUC\_JP のデータベースが存在していることがわかります。

データベース一覧					
名前	所有者	エンコーディング	照合順序	Ctype(変換演算子)	アクセス権限
mydb	fsepta	EUC_JP	C	C	
postgres	fsepta	EUC_JP	C	C	
template0	fsepta	EUC_JP	C	C	=c/fsepta +  fsepta=CTc/fsepta
template1	fsepta	EUC_JP	C	C	=c/fsepta +  fsepta=CTc/fsepta
(4 行)					

- データベースの設定ファイル postgresql.conf ファイルを参照して client\_encoding パラメーターの値を確認

PostgreSQL 文書の「23.3 文字セットサポート」を参照して、上記の \l コマンドで確認したデータベースのエンコーディングとの組み合わせが自動文字セット変換でサポートされているかを確認します。この例では、データベース側 : EUC\_JP に対して、クライアント側 : shift\_jis\_2004 はサポートされていません。

```
client_encoding = shift_jis_2004
```

- client\_encoding パラメーターの値を、実際のクライアントのエンコーディング（SJIS）に書き換える

- 備考 EUC\_JP と組み合わせ可能な日本語の文字セットは、本ページ内の「ポイント」に掲載している組み合わせ表を参照してください。

```
client_encoding = sjis
```

4. 設定ファイルを再読み込みして、変更を反映

```
pg_ctl reload
```

5. client\_encoding パラメーターに誤りがなかった場合、クライアント側の PGCLIENTENCODING 環境変数に設定されている文字セットに誤りがある可能性があります。OS のコマンドを使用して PGCLIENTENCODING 環境変数の設定値を確認し、自動文字セット変換でサポートされている文字セットに変更してください。

## ポイント

- PostgreSQL では、initdb コマンドによるデータベースクラスタ初期化の際に、データベースのデフォルトの文字セットを指定できます。さらに、データベース作成時にデフォルトとは異なる文字セットを指定することで、同じデータベースクラスタ内で異なる文字セットのデータベースを使用することができます。このデータベース側の文字セットに対して、クライアント側に設定された文字セットが異なる場合、PostgreSQL はサポートする組み合わせの範囲で、自動的に文字セットを変換します。サポート外の組み合わせが指定された場合は、以下のようなエラーが output されます。

### エラーメッセージの例

initdb コマンドの -E または --encoding オプションに EUC\_JP を指定して初期化します。生成された postgresql.conf ファイルで client\_encoding に SHIFT\_JIS\_2004 を設定し、pg\_ctl コマンドでデータベースサーバーを起動すると、SHIFT\_JIS\_2004 と EUC\_JP の組み合わせがサポートされていないため、以下のメッセージが出力されます。

```
FATAL: SHIFT_JIS_2004 と EUC_JP 間の変換はサポートされていません
```

- データベースサーバーとの接続時にクライアント側のデフォルトの文字セットを設定するためには、以下の方法があります。
  - postgresql.conf ファイルの client\_encoding パラメーターを設定
  - 個々のクライアントにおいて PGCLIENTENCODING 環境変数を設定
- 日本語の文字セットにおいて、自動文字セット変換がサポートする組み合わせを、以下に示します。

【凡例】 レ : サポート、N/A : 未サポート

データベース	クライアント				
	EUC_JP	EUC_JIS_2004	SJIS	SHIFT_JIS_2004	UTF-8
EUC_JP	レ	N/A	レ	N/A	レ
EUC_JIS_2004	N/A	レ	N/A	レ	レ
UTF-8	レ	レ	レ	レ	レ

- 備考 SJIS と SHIFT\_JIS\_2004 はデータベース側では使用できません。

- データベース側とクライアント側の間で自動文字セット変換が実行されると性能上のロスが発生するため、可能であれば同一の文字セットを使用することをお勧めします。
- 自動文字セット変換は、PostgreSQL の pg\_conversion システムカタログに予め格納された変換情報の範囲で実施されます。
- 実際に扱うデータの文字セットが自動文字セット変換でサポートされていない組み合わせの場合は、SQL 文の CREATE CONVERSION コマンドを利用して新規の自動文字セット変換を追加するか、または、予めデータの文字セット自体を変換しておく必要があります。

## 参考

### PostgreSQL 13.1 文書

- Documentation (PostgreSQL オフィシャル)  
<https://www.postgresql.org/docs/>
  - III. Server Administration
    - 23. Localization
      - 23.3 Character Set Support
- PostgreSQL 日本語ドキュメント (日本 PostgreSQL ユーザ会)  
<https://www.postgresql.jp/document/>
  - III. サーバの管理
    - 23. 多言語対応
      - 23.3 文字セットサポート

2021 年 7 月 30 日