



管理：グローバルゼーション

---

**SAP Sybase IQ 16.0 SP03**

ドキュメント ID：DC02003-01-1603-01

改訂：2013 年 11 月

Copyright © 2013 by SAP AG or an SAP affiliate company. All rights reserved.

このマニュアルの内容を SAP AG による明示的な許可なく複製または転載することは、形態や目的を問わず禁じられています。ここに記載された情報は事前の通知なしに変更されることがあります。

SAP AG およびディストリビュータが販売しているソフトウェア製品には、他のソフトウェアベンダ独自のソフトウェアコンポーネントが含まれているものがあります。国内製品の仕様は変わることがあります。

これらの資料は SAP AG および関連会社 (SAP グループ) が情報のみを目的として提供するものであり、いかなる種類の表明または保証も行わないものではなく、SAP グループはこの資料に関する誤りまたは脱落について責任を負わないものとします。SAP グループの製品およびサービスに関する保証は、かかる製品およびサービスに付属している明確な保証文書がある場合、そこで明記されている保証に限定されます。ここに記載されているいかなる内容も、追加保証を構成するものとして解釈されるものではありません。

ここに記載された SAP および他の SAP 製品とサービス、ならびに対応するロゴは、ドイツおよび他の国における SAP AG の商標または登録商標です。その他の商標に関する情報および通知については、<http://www.sap.com/corporate-en/legal/copyright/index.epx#trademark> を参照してください。

# 目次

国際言語サポートについて .....	1
各国語と文字セット .....	1
ICU とは何か、いつ必要になるか .....	2
文字セット .....	4
Windows でのコードページ .....	4
照合を使用した文字のソート .....	8
照合 .....	9
SAP Sybase IQ 照合アルゴリズム (SACA) .....	9
Unicode 照合アルゴリズム (UCA) .....	10
SAP Sybase IQ データベース内での照合 .....	12
代替照合 .....	13
トルコ語文字セットと照合 .....	16
ロケール .....	19
ロケール言語 .....	19
ロケール文字セット .....	21
文字セット変換について .....	22
データベースメッセージの文字変換 .....	22
接続文字列と文字セット .....	23
文字セット変換の回避 .....	23
<b>ANSI コードページと OEM コードページ .....</b>	<b>25</b>
ANSI ISO_1 照合 .....	25
ANSI 1252LATIN1 照合 .....	25
ANSI ISO1LATIN1 照合 .....	26
ANSI ISO9LATIN1 照合 .....	26
<b>照合の表示 .....</b>	<b>27</b>
<b>マルチバイト照合 .....</b>	<b>29</b>
日本語のサポート .....	29
タイ語のサポート .....	29

ロケール情報 .....	31
特定のロケールをサポートする照合の選択 .....	33
文字セットと照合順の選択によるパフォーマンスの最大 化 .....	35
ロケールの設定 .....	37
言語ラベル値のリスト .....	37
INSERT...LOCATION 文のロケールの設定 .....	38
データベースサーバで文字セット変換を無効にする .....	41
デフォルトの照合を使用したデータベースの作成 .....	43
データベースの照合の変更 .....	45
索引 .....	47

# 国際言語サポートについて

データベースを作成するときは、そのデータベースで使用する照合順(「照合」とも呼ぶ)を指定します。照合とは、「文字セット」と、データベース内の文字の「ソート順」を組み合わせたものです。

## 各国語と文字セット

---

SAP® Sybase® IQ インストール環境で各国語を扱えるように設定します。

データベース照合では、オブジェクト名(テーブル名やカラム名など)をはじめとする、データベース内のすべての文字データ型がソートされ、比較されます。

SAP Sybase IQ では、SAP Sybase SQL Anywhere® Collation Algorithm の容量効率と速度を利用しています。

- データベースオプション `SORT_COLLATION` を使用すると、**ORDER BY** 式で **SORTKEY** 関数を暗黙的に使用できます。このオプションの値を有効な照合名または照合 ID に設定すると、**ORDER BY** 句の文字列式は、**SORTKEY** 関数が呼び出されたものとして扱われます。
- **SORTKEY** 関数は、Sybase Unicode Infrastructure Library (Unilib®) の代わりに、International Components for Unicode (ICU) ライブラリを使用します。15 より前のバージョンの SAP Sybase IQ を使用して作成されたソートキー値には、それ以降のバージョンを使用して作成された値と同じ値は含まれていません。15 より前のバージョンの SAP Sybase IQ を使用して生成されたデータベース内のソートキー値は、すべて再生成してください。
- **CREATE DATABASE** のパラメータ **COLLATION** では、データベースに対する照合の指定がサポートされています。  
データベースの照合は、オペレーティングシステムで使用される照合と一致する必要があります。**CREATE DATABASE** のデフォルトの文字セットは `ISO_BINENG` です。
- **CP874toUTF8** ユーティリティは、CP874 文字セットのデータを UTF8 照合(タイ語用に SAP Sybase IQ がサポートしている照合)に変換します。**CP874toUTF8** ユーティリティは、ICU ライブラリを呼び出してデータ変換を実行します。このユーティリティを使用して、データを UTF8 に変換せずに CP874 文字セットのデータをロードすることもできます。

SAP Sybase IQ では、カスタム照合はサポートされなくなっています。カスタム照合の使用を再構築する場合、カスタム照合を保持するには 1 ステップで再構築してください。データベースをアンロードしてから、作成したデータベースにス

キーマとデータをロードする場合は、用意されているいずれかの照合を使用する必要があります。

データベース照合の変更点と、SAP Sybase IQ 15 で廃止された照合のリストについては、『SAP Sybase IQ 15.0 の新機能』を参照してください。

廃止された照合を使用して作成された既存の 12.7 データベースから最新バージョンの SAP Sybase IQ に移行するには、**iqunload** ユーティリティを使用します。

## ICU とは何か、いつ必要になるか

ICU (International Components for Unicode) は、IBM が開発および保守しているオープンソースライブラリです。ICU は、Unicode サポートを提供することによって、ソフトウェアの国際化を容易にします。SAP Sybase IQ は、ICU を使用して、特定の文字セット変換と照合オペレーションを実装しています。

データベースサーバで ICU が必要になるとき (*Windows Mobile* を除くすべてのプラットフォーム)

データベースサーバで常に ICU を利用できることが理想的です。次の表に、ICU が必要になるときとその理由を示します。

ICU が必要になるとき	説明
NCHAR 文字セットまたは CHAR 文字セットの照合として Unicode Collation Algorithm (UCA) が使用されている	UCA は ICU を必要とします。
データベースの文字セットは UTF-8 でなく、マルチバイト文字セットである	データベースの文字セットから UTF-8 にパスワード変換するために必要です (データベースパスワードは内部的に UTF-8 で格納されています)。
クライアントとデータベースの文字セットが異なり、いずれかがマルチバイト (UTF-8 を含む) である。これには、Unicode ODBC、OLE DB、ADO.NET、SAP Sybase IQ JDBC アプリケーションが含まれます。ICU のインストールされていないクライアントが使用するデータベース文字セットには関係ありません。	マルチバイト文字セットの適切な変換を行うためには ICU が必要です。
データベースの文字セットが UTF-8 でなく、CHAR と NCHAR 間の変換が必要である	データベースサーバは、UTF-8 を別の文字セットに変換するために ICU を必要とします。

ICU が必要になるとき	説明
Embedded SQL クライアントが UTF-8 以外の NCHAR 文字セットを使用する	データベースサーバは、UTF-8 を別の文字セットに変換するために ICU を必要とします。Embedded SQL クライアントのデフォルトの NCHAR 文字セットは、最初のクライアントの CHAR 文字セットと同じです。これは、 <code>db_change_nchar_charset</code> 関数を使用して変更できます。
CSCONVERT または SORTKEY 関数が使用されている。CSCONVERT 関数は、上記の 3 番目の項目要件に準ずる文字セットを変換するために呼び出されます。	マルチバイト文字セットの文字セット変換を行うためには ICU が必要です。SORTKEY のラベル用に SORTKEY を生成するには UCA が必要であるため、結果として ICU が必要になります。

データベースサーバ上で ICU を使用しないで正しい文字セット変換を実行できる  
とき

データベースの文字セットとクライアントの文字セットの両方がシングルバイトであり、`sqlany.cvf` を利用できる時 (すべてのプラットフォーム)、またはオペレーティングシステムが変換をサポートするとき (Windows のみ) は、ICU を使用しないで正しい文字セット変換を行うことができます。これは、`sqlany.cvf` を利用できる場合、またはホストのオペレーティングシステムに適切な変換機能がインストールされている場合にかぎり、ICU を使用しなくてもシングルバイト間の変換が処理可能なためです。

クライアントで ICU が必要になるとき (すべてのプラットフォーム)

Unicode のクライアントアプリケーションでは、使用するデータベースの文字セットにかかわらず、すべてのクライアントに ICU がインストールされている場合に、クライアントとデータベースサーバ間のパフォーマンスが向上する可能性があります。これは、必要な変換処理の一部がデータベースサーバからクライアントに渡されて、負荷が軽減するからです。

また、Windows プラットフォームで ODBC を使用している場合は、ANSI アプリケーションであっても、クライアントに ICU をインストールしておくことが必要です。これは、ドライバマネージャが ANSI ODBC 呼び出しを Unicode ODBC 呼び出しに変換するからです。

## 文字セット

---

各ソフトウェアは、**文字セット**を使用します。文字セットは記号、文字、数字、スペースなどから成ります。"ISO-8859-1"は文字セットの例です。Latin1とも呼ばれます。

文字を内部的に適切に表すため、各ソフトウェアは**エンコード** (**文字コード**とも呼ばれる)を使用します。エンコードとは、各文字を1バイトまたは複数バイトの情報にマッピングする方法で、16進数で表します。UTF-8はエンコードの例です。

「文字セット」と「エンコード」は密接に関連しており、どちらも「エンコード」の意味で使用されることがあります。

**コードページ**は、エンコードの一形態です。コードページとは文字と数値表現とのマッピングのことで、通常、数値表現は0～255の整数です。コードページの例には、Windows コードページ 1252があります。

このマニュアルでは、「エンコード」、「文字コード」、「文字セットエンコード」、「コードページ」を同じ意味で使用します。

データベースサーバは、文字をソート (たとえば、名前をアルファベット順にリスト) するときに**照合**を使用します。照合は文字コード (文字と表現間のマッピング) と文字の**ソート順**の組み合わせです。各文字列にソート順が複数ある場合があります。たとえば、大文字/小文字を区別するソート順と大文字/小文字を区別しないソート順があります。また、言語間で同じ文字に対するソート順が異なる場合もあります。

文字は**フォント**を使って画面上に表示されます。これは文字セットの文字とその外観との間のマッピングです。フォントはオペレーティングシステムによって処理されます。

オペレーティングシステムは、**キーボードマッピング**を使って、キーボードのキーまたはキーの組み合わせを文字セットの文字にマッピングします。

## Windows でのコードページ

---

多くの言語では、文字の数はシングルバイト文字セットで扱える程度です。シングルバイト文字セットでは、各文字が1「バイト」(2桁の16進数)で表されます。

シングルバイトでは最大256文字を表すことができます。アクセント記号の付いた文字を含め、世界で使用されているすべての文字を1つのシングルバイト文字セットで表現することはできません。この問題は、1つまたは複数の言語に適した文字セットをそれぞれ記述した**コードページ**の使用により解決します。たとえば、コードページ 869はギリシャ語の文字セット、コードページ 850は各国語でさまざまな文字を表現するのに適した国際的な文字セットになっています。



**Windows 環境の ANSI コードページと OEM コードページ**

Windows の場合、2つのコードページが使用されています。Windows グラフィカルユーザインタフェースを使用するアプリケーションでは、Windows コードページが使用されます。Windows コードページには、ISO 文字セットおよび ANSI 文字セットとの互換性があります。このようなコードページは、しばしば **ANSI コードページ**と呼ばれます。

Windows で動作する文字モードアプリケーション(コマンドプロンプトウィンドウを使用するアプリケーション)は DOS で使用されていたコードページを使用します。これは歴史的な理由から **OEM コードページ** (Original Equipment Manufacturer) と呼ばれます。

SAP Sybase IQ は、OEM と ANSI コードページの両方に基づいた照合をサポートします。OEM 照合は互換性確保のためにサポートされていますが、新しいデータベースでは使用しないでください。

**推奨文字セットと照合**

SAP Sybase IQ は数多くの文字セット、コードページ、エンコード、照合の名前を認識しますが、このトピックでは Windows プラットフォームと UNIX プラットフォームで使用が推奨されるものを使用言語別に示します。

iqinit -le オプションを使用すると、SAP Sybase IQ データベースで使用可能なすべての文字セットラベルのリストを取得できます。iqinit -l オプションを使用すると、SAP Sybase IQ データベースで使用可能な照合のリストを取得できます。

文字セットエンコードまたはラベルを指定する必要がある場合は、「文字セットラベル」列の値または iqinit -le によって表示されるいずれかのラベルを使用します。照合を指定する必要がある場合は、「照合」列か「代替照合」列の値または iqinit -l によって表示されるいずれかのラベルを使用します。

**注意：** 次の表に示されていない言語については、UTF-8 エンコードを UCA 照合または UTF8BIN 照合と組み合わせて使用してください。

**Windows プラットフォーム**

言語	Windows コードページ	文字セットラベル	照合	代替照合
アラビア語	1256	Windows -1256	1256ARA	
中央および西ヨーロッパ言語	1250	Windows -1250	1250LATIN2	

国際言語サポートについて

言語	Windows コードペー ジ	文字セットラベ ル	照合	代替照合
デンマーク語	1252	Windows-1252	1252LATIN1	
オランダ語	1252	Windows-1252	1252LATIN1	
英語	1252	Windows-1252	1252LATIN1	
フィンランド語	1252	Windows-1252	1252SWEFIN	
フランス語	1252	Windows-1252	1252LATIN1	
ドイツ語	1252	Windows-1252	1252LATIN1	
ギリシャ語	1253	Windows -1253	1253ELL	
ヘブライ語	1253	Windows-1253	1255HEB	
イタリア語	1252	Windows-1252	1252LATIN1	
日本語	932	Windows-31J	932JPN	
韓国語	949	IBM949	949KOR	
リトアニア語	1257	Windows -1257	1257LIT	
ノルウェー語	1252	Windows-1252	1252NOR	
ポーランド語	1250	Windows-1250	1250POL	
ポルトガル語	1252	Windows-1252	1252LATIN1	
ロシア語	1251	Windows -1251	1251CYR	
中国語 (簡体文字)	936	GBK	936ZHO	
スペイン語	1252	Windows-1252	1252SPA	
スウェーデン語	1252	Windows-1252	1252SWEFIN	
タイ語	874	TIS-620	874THAIBIN	
中国語 (繁体文字) - 香 港	950	Big5-HKSCS	950ZHO_HK	
中国語 (繁体文字) - 台 湾	950	Big5	950ZHO_TW	
トルコ語	1254	Windows -1254	1254TRK	1254TRKALT
ウクライナ語	1251	Windows-1251	1251CYR	

言語	Windows コードページ	文字セットラベル	照合	代替照合
西ヨーロッパ言語	1252	Windows-1252	1252LATIN1	

## UNIX プラットフォーム

言語	文字セットラベル	照合	代替照合
アラビア語	ISO_8859-6:1987	UCA	
中央および西ヨーロッパ言語	ISO_8859-2:1987	UCA	
デンマーク語	ISO-8859-15	ISO9LATIN1	ISO1LATIN1
オランダ語	ISO-8859-15	ISO9LATIN1	ISO1LATIN1
英語	ISO-8859-15	ISO9LATIN1	ISO1LATIN1
フィンランド語	ISO-8859-15	ISO9LATIN1	ISO1LATIN1
フランス語	ISO-8859-15	ISO9LATIN1	ISO1LATIN1
ドイツ語	ISO-8859-15	ISO9LATIN1	ISO1LATIN1
ギリシャ語	ISO_8859-7:1987	UCA	
ヘブライ語	ISO_8859-8:1988	UCA	
イタリア語	ISO-8859-15	ISO9LATIN1	ISO1LATIN1
日本語	EUC-JP <sup>1</sup>	EUC_JAPAN	
韓国語	EUC-KR	EUC_KOREA	
リトアニア語	(UTF-8 を使用)	UCA または UTF8BIN	
ノルウェー語	ISO-8859-15	ISO9LATIN1	ISO1LATIN1
ポーランド語	ISO_8859-2:1987	UCA	
ポルトガル語	ISO-8859-15	ISO9LATIN1	ISO1LATIN1
ロシア語	ISO_8859-5:1988	UCA	
中国語(簡体文字)	GB2312	EUC_CHINA	
スペイン語	ISO-8859-15	ISO9LATIN1	ISO1LATIN1
スウェーデン語	ISO-8859-15	ISO9LATIN1	ISO1LATIN1
タイ語	(UTF-8 を使用)	UCA または UTF8BIN	

言語	文字セットラベル	照合	代替照合
中国語 (繁体文字) - 香港	Big5-HKSCS	950ZHO_HK	950TWN
中国語 (繁体文字) - 台湾	EUC-TW	EUC_TAIWAN	
中国語 (繁体文字) - 台湾	Big5	950ZHO_TW	
トルコ語	ISO_8859-9:1989	920TRK	
ウクライナ語	ISO_8859-5:1988	UCA	
西ヨーロッパ言語	ISO-8859-15	ISO9LATIN1	ISO1LATIN1

<sup>1</sup> EUC-JP は、SAP Sybase IQ ラベルである  
Extended\_UNIX\_Code\_Packed\_Format\_for\_Japanese の代替ラベルです。

## 照合を使用した文字のソート

データベースの照合順は、文字のアルファベット順をもとに、その範囲を広げて数字やスペースを含めた文字セットのすべての文字に対応したものです。

各ソート位置には複数の文字を割り当てることができます。これは、たとえばアクセント付きの文字をアクセントなしの文字と同じに扱うときなどに便利です。

ソート位置が同じ 2 つの文字はデータベースではまったく同じものとして扱われます。したがって、照合で 'a' と 'e' が同じソート位置に割り当てられている場合、次のような結果になります。

```
WHERE coll = 'want'
```

上のような検索条件を持つクエリでは、**col1** のエントリが **went** であるローも条件を満たします。

ソート位置ごとに大文字と小文字を指定できます。大文字と小文字を区別するデータベース (SAP Sybase IQ データベースではこれがデフォルト) では、大文字と小文字が別のもんとして扱われます。区別されないデータベースでは、同じものとして扱われます。

**注意：** ドイツ語のシステムのデフォルト照合を選択するコードでは、1252DEU ではなく 1252LATIN1 を選択してください。1252DEU ではウムラウトのある文字とない文字が区別されるのに対し、1252LATIN1 では区別されません。1252LATIN1 では Muller と Müller は同じとみなされますが、1252DEU では違くとみなされません。1252DEU ではウムラウトが付いた文字は別の文字と判断されるため、ob、öa、というアルファベット順になります。

### マルチバイト文字セットの最初のバイトによる照合順

マルチバイト文字のソート順は、最初のバイトにのみ設定できます。

最初のバイトの値が同じ文字は、それ以降のバイトの 16 進値によってソートされます。

## 照合

照合とは、特定の文字セットまたはエンコードに対する文字のソートおよび比較の方法です。SAP Sybase IQ では、照合アルゴリズムとして、SAP Sybase IQ 照合アルゴリズム (SACA) と Unicode 照合アルゴリズム (UCA) の 2 つをサポートしています。SACA を使用すると、ソートが高速、簡潔、実用的になりますが、言語的な正確さが低下します。UCA を使用すると、言語的な処理は正確になりますが、記憶領域の要件と実行時間が多少増加します。

### SAP Sybase IQ 照合アルゴリズム (SACA)

SAP Sybase IQ 照合アルゴリズム (SACA) は、シングルバイト文字とマルチバイト文字の実用的な比較、ソート、大文字小文字変換を提供します。このアルゴリズムは、必要とする領域が少なく、高速です。インデックスなどの文字列は、マッピング後も元の文字列と長さが同じです。比較、ソート、大文字小文字変換のマッピングでは、文字列の各バイト値ごとの簡単なテーブルルックアップを使用しています。

#### シングルバイト文字セット

典型的なシングルバイト文字セットの照合では、各文字のアクセント記号付きとアクセント記号なしのあらゆる形が同じ値にマッピングされており、照合ではアクセント記号は区別されません。同じ文字のアクセント記号付き形とアクセント記号なし形は同じとみなされ、ソートされると隣り合わせになります。

この照合では、アクセント記号を維持したまま、大文字と小文字が変換されます。

#### マルチバイト文字セット

マルチバイト文字セットの場合、リードバイトが 256 個の異なる値にマッピングされています。フォロワーバイトはバイナリ値として比較されます。

マルチバイト文字セットのほとんどの照合では、文字セットエンコードにより、文字はリードバイトによって識別される 256 バイト単位のページにグループ化されるので、このマッピング方法で実用的なソート結果が得られます。これらのページと各ページに含まれる文字は、該当する文字セットとして合理的な順序で配置されています。通常、この照合では文字セットにおけるページの順序 (リードバイト) が保持されます。ページによっては、他の特徴に基づいてソートされま

す。たとえば、日本語コードページ 932 用の 932JPN 照合では、全角文字 (漢字) と半角文字 (カタカナ) をグループ分けしています。

大文字小文字変換は、7 ビットの英文字についてのみ提供されています。

### UTF-8 文字セット

UTF-8 はマルチバイト文字セットです。各文字は 1 ~ 4 バイトで構成されます。SAP Sybase IQ では、UTF-8 文字のソート用に UTF8BIN 照合を提供しています。

UTF8BIN では、リードバイトが 256 の異なる値にマッピングされており、フォロワーバイトはバイナリ値として比較されます。UTF-8 の文字表現方法と 256 というマッピングの数的制限に起因して、同じ文字のアクセント記号付き形とアクセント記号なし形など、関連文字をグループ化できません。ソートの基準は基本的にバイナリです。

大文字小文字変換は、7 ビットの英文字についてのみサポートされています。

## Unicode 照合アルゴリズム (UCA)

Unicode 照合アルゴリズム (UCA) は、Unicode 文字セット全体のソートに使用するアルゴリズムです。これにより、言語的に正しい比較、ソート、大文字小文字変換が実現されます。UCA は Unicode 規格の一部として開発されました。SAP Sybase IQ では、IBM が開発および保守している International Components for Unicode (ICU) オープンソースライブラリを使用して UCA を実装しています。

---

**注意：** デフォルトの UCA ソート順により、ほとんどの言語のほとんどの文字が適切な順序でソートされます。ただし、同じ文字を使用する言語間でソートや比較にさまざまな違いがあるため、UCA ですべての言語について適切なソート順が得られるわけではありません。そのため、ICU は UCA を調整できる構文を提供しています。

---

UCA を使用すると、少ない領域と時間で高度な比較、ソート、大文字小文字変換を実現できます。

マッピング後の文字列は元の文字列より長くなります。このアルゴリズムは、複雑な文字を的確に処理できます。

SAP Sybase IQ 照合アルゴリズム (SACA) とは異なり、Unicode 照合アルゴリズム (UCA) はシングルバイトの UTF-8 文字セットにのみ使用され、各文字を 1 つまたは複数の属性で区別します。文字の場合、属性は基底文字、アクセント記号、大文字小文字です。

文字以外の場合、通常は基底文字だけが属性になります。

UCA は、次の方法で文字を比較します。

- 基底文字を比較します。文字列の基底文字が他の文字列と異なる場合は、その時点で比較が完了します。アクセント記号や大文字小文字の違いは考慮されません。
- データベースでアクセント記号の違いが区別される場合は、アクセント記号が比較されます。アクセント記号が異なる場合、その時点で比較が完了します。大文字小文字の違いは考慮されません。
- データベースで大文字と小文字が区別される場合は、各文字の大文字と小文字が比較されます。

元の文字列の値が同じとみなされるのは、基底文字、アクセント記号、大文字小文字がまったく同じ場合だけです。

### 例

UCA を使用して次の表の第 1 カラムの文字列を比較するとします。後続のカラムには各文字列の 3 つの属性が記載されています。基底文字は同じで、違うのはアクセント記号と大文字小文字だけです。

文字列	基底文字	アクセント記号	大文字小文字
noel	noel	なし、なし、なし、なし	小、小、小、小
noël	noel	なし、なし、アクセント記号、なし、	小、小、小、小
Noel	noel	なし、なし、なし、なし	大、小、小、小
Noël	noel	なし、なし、アクセント記号、なし、	大、小、小、小

次の表は、UCA を使用した場合にアクセント記号と大文字小文字の区別により可能な 4 つの条件によるソート順を示します。

アクセント記号の区別	大文字と小文字の区別	ORDER BY による結果	説明
N	N	Noel、noël、Noël、noel (順不同)	<ul style="list-style-type: none"> <li>• アクセント記号は区別しない</li> <li>• 大文字と小文字も区別しない</li> <li>• すべての値が同じとみなされる</li> <li>• 順序は要素数 4 の集合内でランダム</li> </ul>

アクセント記号の 区別	大文字と 小文字の 区別	ORDER BY による 結果	説明
Y	N	Noel, noel (順不同)、 その後に noël, Noël (順不同)	<ul style="list-style-type: none"> <li>アクセント記号なしが先、アクセント記号ありが後。したがってëの前にeがくる</li> <li>大文字と小文字は区別しない。Nとnは2つの間でランダム</li> </ul>
N	Y	Noel, Noël (順不同)、 その後に noël, noel (順不同)	<ul style="list-style-type: none"> <li>大文字が先、小文字が後。したがってnの前にNがくる</li> <li>アクセント記号は区別しない。eとëは2つの間でランダム</li> </ul>
Y	Y	Noel noel Noël noël	<ul style="list-style-type: none"> <li>アクセント記号なしが先、アクセント記号ありが後。したがってëの前にeがくる</li> <li>大文字が先、小文字が後。したがってnの前にNがくる</li> </ul>

## SAP Sybase IQ データベース内での照合

### CHAR 照合

CHAR データ型 (CHAR、VARCHAR、LONG VARCHAR など) の照合では、SAP Sybase IQ 照合アルゴリズム (SACA) が使用される場合と Unicode 照合アルゴリズム (UCA) が使用される場合があります。どちらの場合も、使用される照合は CHAR 照合と呼ばれます。

### NCHAR 照合

NCHAR データ型 (NCHAR、NVARCHAR、LONG NVARCHAR など) の照合では、Unicode 照合アルゴリズム (UCA) が使用される場合と UTF8BIN 照合が使用される場合があります。UTF8BIN 照合では SAP Sybase IQ 照合アルゴリズム (SACA) が使用されます。

### 大文字小文字とアクセント記号の区別

作成時に大文字小文字を区別するよう指定されなかった SAP Sybase IQ データベースでは、大文字と小文字は区別されません。区別されるようにするには、該当するオプションを指定します。データベースの作成後、データベースを再構築しないで大文字と小文字が区別されるようにすることはできません。



データベースが大文字と小文字を区別するかどうかによって、SACA 照合でも UCA 照合でも大文字と小文字を区別するかどうかが決まり、それにより CHAR 照合と NCHAR 照合で大文字と小文字を区別するかどうかが決まります。

作成時にアクセント記号を区別するよう指定されなかった SAP Sybase IQ データベースでは、アクセント記号は区別されません。区別されるようにするには、該当するオプションを指定します。データベースの作成後、データベースを再構築しないでアクセント記号が区別されるようにすることはできません。

データベースがアクセント記号を区別するかどうかは、UCA 照合にのみ影響を与えます。UCA 照合が CHAR 照合や NCHAR 照合で使用されているかどうかには関係ありません。CHAR 照合と NCHAR 照合のどちらにも SACA 照合を使用することを選択した場合、アクセント記号に関するオプションに効力はありません。アクセント記号の区別は SACA 照合の属性であり、該当するオプションをデータベース作成時に使用して指定することはできません。

## 代替照合

次の表に、SORTKEY および COMPARE 関数で利用できる互換性の照合を示します。

照合ラベル	説明
874THAIBIN	Code Page 874、Windows Thai、ISO8859-11、バイナリ順序
932JPN	Code Page 932、日本語シフト JIS、Microsoft 拡張文字付き
936ZHO	Code Page 936、中国語 (簡体文字)、PRC GBK 2312-80 8 ビットコード
949KOR	Code Page 949、韓国語 KS C 5601-1987 コード、完成型
950ZHO_HK	Code Page 950、中国語 (繁体文字)、Big 5 コード (HKSCS を含む)
950ZHO_TW	Code Page 950、中国語 (繁体文字)、Big 5 コード
1250LATIN2	Code Page 1250、Windows Latin 2、中央/東ヨーロッパ言語
1250POL	Code Page 1250、Windows Latin 2、ポーランド語
1251CYR	Code Page 1251、キリル語
1252BIN	コードページ 1252、Windows ラテン語 1、バイナリ順序
1252LATIN1	Code Page 1252、Windows Latin 1、西ヨーロッパ言語
1252NOR	Code Page 1252、Windows Latin 1、ノルウェー語
1252SPA	Code Page 1252、Windows Latin 1、スペイン語

照合ラベル	説明
1252SWEFIN	Code Page 1252、Windows Latin 1、スウェーデン語/フィンランド語
1253ELL	Code Page 1253、Windows ギリシア語、ISO8859-7 拡張付き
1254TRK	Code Page 1254、Windows Latin 5、トルコ語、ISO 8859-9 拡張付き
1254TRKALT	Code Page 1254、Windows トルコ語、拡張付き ISO8859-9、I-dot と I-no-dot は同じ
1255HEB	Code Page 1255、Windows ヘブライ語、ISO8859-8 拡張付き
1256ARA	Code Page 1256、Windows アラビア語、ISO8859-6 拡張付き
1257LIT	Code Page 1257、リトアニア語
CESU8BIN	CESU-8、Unicode 用の 8 ビットマルチバイトエンコード、バイナリ順序
EUC_CHINA	中国語(簡体文字)の GB 2312-80 コード
EUC_JAPAN	日本語の EUC JIS X 0208-1990 と JIS X 0212-1990 コード
EUC_KOREA	韓国語の KS C 5601-1992 コード、Johab (組合型)
EUC_TAIWAN	台湾語の Big 5 コード
ISO1LATIN1	ISO8859-1、ISO Latin 1、西ヨーロッパ言語、Latin 1 順序
ISO9LATIN1	ISO8859-15、ISO Latin 9、西ヨーロッパ言語、Latin 1 順序
ISO_1	ISO8859-1、Latin 1、西ヨーロッパ言語
ISO_BINENG	バイナリ順序、英語 ISO/ASCII 7 ビット文字ケースマッピング
UCA	標準のデフォルト UCA 照合
UTF8BIN	UTF-8、Unicode 用 8 ビットマルチバイトエンコード、バイナリ順序

#### 代替照合

代替照合は、古いバージョンの SAP Sybase IQ との互換性や、その他の特殊な用途で使用できます。サポートされている代替照合の一覧を表示するには、次のコマンドを実行します。

```
iqinit -l+
```

#### Adaptive Server Enterprise の照合

次の表は、SORTKEY 関数などの機能で使用するためにサポートされている Adaptive Server Enterprise 照合を示します。

説明	照合名	照合 ID
デフォルト Unicode マルチ言語	default	0
CP 850 代替言語：アクセント記号なし	altnoacc	39
CP 850 代替言語：小文字優先	altdict	45
CP 850 西ヨーロッパ言語：大文字／小文字の区別なし、優先	altnocsp	46
CP 850 スカンジナビア語辞書順	scandict	47
CP 850 スカンジナビア語：大文字／小文字の区別なし、優先	scannocp	48
GB ピンイン	gbpinyin	なし
バイナリソート	binary	50
Latin-1 英語、フランス語、ドイツ語辞書順	dict	51
Latin-1 英語、フランス語、ドイツ語、大文字／小文字の区別なし	nocase	52
Latin-1 英語、フランス語、ドイツ語、大文字／小文字の区別、優先設定なし	nocasep	53
Latin-1 英語、フランス語、ドイツ語、アクセント記号なし	noaccent	54
Latin-1 スペイン語辞書順	espdict	55
Latin-1 スペイン語、大文字／小文字の区別なし	espnocs	56
Latin-1 スペイン語、アクセント記号なし	espnoac	57
ISO 8859-5 ロシア語辞書順	rusdict	58
ISO 8859-5 ロシア語、大文字／小文字の区別なし	rusnocs	59
ISO 8859-5 キリル語辞書順	cyrdict	63
ISO 8859-5 キリル語、大文字／小文字の区別なし	cyrnocs	64
ISO 8859-7 ギリシャ語辞書順	elldict	65
ISO 8859-2 ハンガリー語辞書順	hundict	69
ISO 8859-2 ハンガリー語、アクセント記号なし	hunnoac	70
ISO 8859-2 ハンガリー語、大文字／小文字の区別なし	hunnocs	71
ISO 8859-5 トルコ語辞書順	turdict	72

説明	照合名	照合 ID
ISO 8859-5 トルコ語、アクセント記号なし	turnoac	73
ISO 8859-5 トルコ語、大文字／小文字の区別なし	turnocs	74
CP 874 (TIS 620) タイ語辞書順	thaidict	1
ISO 14651 順序付け標準	14651	22
シフト JIS バイナリ順	sjisbin	179
Unicode UTF-8 バイナリソート	utf8bin	24
EUC JIS バイナリ順	ejisbin	192
GB2312 バイナリ順	gb2312bn	137
CP932 MS バイナリ順	cp932bin	129
Big5 バイナリ順	big5bin	194
EUC KSC バイナリ順	euckscbn	161

## トルコ語文字セットと照合

トルコ語には、I に相当する文字に 2 通りの表記形式があります。1 つ目の形式は、I-dot と呼ばれるもので、次のように表記されます。

ı, İ

2 つ目の形式は、I-no-dot と呼ばれるもので、次のように表記されます。

ı, İ

これらの文字は同じ文字の変形として表示される場合でも、トルコ語のアルファベットでは別の文字と見なされます。SAP Sybase IQ では、トルコ語照合 1254TRK を提供して、このような違いに対処しています。

これらの文字の大文字と小文字の変換に関するトルコ語の規則は、ANSI SQL 標準規則と互換性がありません。たとえば、トルコ語では I の小文字に相当するには、以下の文字です。

ı

一方、ANSI 標準ではこれは次の文字になります。

i

このような理由から、大文字と小文字を区別しない場合に正しく一致させることができるのは、一致させるテキストがトルコ語か英語/ANSI であるかどうかによります。多くのコンテキストでは、これを区別するだけの十分な情報がないので、そのようなデータベースでは標準外の動作となることがあります。

たとえば、次の文を 1254TRK 照合を使用するデータベースに対して実行するとします。

```
SELECT * FROM syshistory //actual table name is SYSHISTORY
SELECT * FROM fig //actual name is FIG
```

最初の文はシステムオブジェクトを参照しており、名前の照合には ANSI SQL 変換規則が必要です。2 番目の文はユーザオブジェクトを参照しており、名前の照合にはトルコ語変換規則が必要です。ここで、データベースサーバは、オブジェクトが何であるかを確認できるまでは使用すべき変換規則を判断できず、使用する変換規則が確認できるまではオブジェクトが何であるかを判断できません。この状況においては、システムおよびユーザオブジェクトの両方を満足させるような解決策はありません。この例では、データベースサーバがトルコ語照合 1254TRK を使用しており、I の小文字が I の大文字と同じであるとは見なされないため、最初の文は失敗します。2 番目の文は成功します。

トルコ語と ANSI 標準には互換性がないため、トルコ語のデータベースに含まれるシステムオブジェクトを参照する場合は、オブジェクト名の大文字と小文字(そのオブジェクトの作成時に使用された大文字と小文字)を正確に指定する必要があります。上記の最初の文は次のように記述します。

```
SELECT * FROM SYSHISTORY
```

実際は、文字 I のみ大文字と小文字を正しく対応させます。

別の方法として、通常のやり方ではありませんが、次のように文を記述する方法も可能です。

```
SELECT * FROM syshistory //I-no-dot
```

INSERT などのキーワードは、トルコ語データベースでも大文字小文字の区別がありません。SAP Sybase IQ では、すべてのキーワードが英語の文字のみを使用していると認識しているため、キーワードの一致に ANSI の大文字と小文字の変換規則を使用します。また SAP Sybase IQ は、組み込み関数などの特定の識別子にこの方法を適用します。ただし、カタログ内に保管されている名前を持つオブジェクトは、上記のように正しい大文字と小文字または文字を使用して指定してください。正しい大文字と小文字を確認するには、システムオブジェクトを定義するシステムビューを調べてください。

### 大文字と小文字を区別しないトルコ語データベースのデータ

同様の規則で、大文字小文字を区別しないトルコ語データベースのデータを管理します。たとえば、データ値が次のような場合、

**FIG**

上記のデータへの小文字参照は、次のようになります。

**fig**

同じ I-dot 文字が両方の形式で使用されます。

### 代替トルコ語照合 1254TRKALT

一部のアプリケーション開発者にとって、トルコ語の文字の問題が重大な問題を引き起こす場合もあります。正しい解決策は、すべてのオブジェクト参照先で大文字と小文字が正しく対応していることを確認するか、適切な文字 I を使用しているかを確認することです。ただし、トルコ語の規則に合わせず ANSI 規則を優先した方がうまくいく場合もあります

SAP Sybase IQ には、照合 1254TRKALT があります。これは、I-dot と I-no-dot を同等の文字とする以外は 1254TRK と同じです。

この変更に伴う違いをきちんと理解しておく必要があります。1254TRKALT データベースでは、次の文字列の区別がありません。

**fig**

**fig**

これはトルコ語ユーザにとって正しくはありませんが、他のユーザにとっては許容範囲である場合もあります。

2 番目の問題は、ORDER BY を使用する場合に出てきます。次の文字列について考えてみます。

**ia**

**ia**

**is**

**is**

1254TRK データベースでは、文字列の ORDER BY は次のような順序を生成しません。

**ia**

**is**

**ia**

**is**

これは、I-no-dot は I-dot より小さいからです。1254TRKALT データベースでは、次のような順序になります。

```
ia  
la  
ls  
is
```

これは、I-no-dot と I-dot が同等だからです。

## ロケール

---

データベースサーバとクライアントライブラリはどちらも、**ロケール定義**を使用して、言語と文字セット環境を認識します。

アプリケーションのロケールまたはクライアントのロケールは、データベースサーバへの要求時にクライアントまたはクライアントライブラリによって使用され、返す結果に使用される文字セットと、エラーメッセージ、警告、その他のメッセージの言語を決定します。データベースサーバは自身のロケールをアプリケーションのロケールと比較し、必要な文字セット変換を判断します。サーバ上のデータベースによって、ロケール定義が異なる場合があります。また、クライアントによってロケールが異なる場合もあります。

ロケールは次のコンポーネントで構成されています。

- **言語** – ISO 639-1 標準の値を使用した 2 文字の文字列です (たとえばドイツ語は DE)。データベースサーバとクライアントのどちらにも、自身のロケールに対する言語の値があります。

データベースサーバは、ロードする言語ライブラリをロケール言語で判断します。データベースの作成時に照合が指定されていない場合、データベースサーバは使用する照合の判断に文字セットと併せてロケール言語も使用します。

クライアントライブラリでは、ロードする言語ライブラリをロケール言語で判断し、データベースからの要求もロケール言語で判断します。

- **文字セット** – 文字セットとは、使用しているコードページまたはエンコードのことです。クライアントとサーバのどちらにも文字セットの値がありますが、両者が異なる場合もあります。異なる場合は、文字セット変換を実行して相互運用性を確保します。

## ロケール言語

---

ロケール言語は、クライアントアプリケーションのユーザによって使用される言語、またはデータベースサーバのユーザによって使用されることが予測される言語です。

クライアントライブラリとデータベースサーバはどちらも同じ方法でロケールの言語コンポーネントを特定します。

## 国際言語サポートについて

1. SALANG 環境変数が指定されている場合は、その値を使用します。
2. Windows では、SALANG 環境変数がない場合、SAP Sybase IQ 言語レジストリのエントリを確認します。
3. オペレーティングシステムの言語設定を確認します。
4. 上記の設定で言語を特定できなかった場合は、デフォルトで英語になります。

### 言語ラベルの値

次の表は、有効な言語ラベルの値と対応する ISO 639 言語コードを示します。

言語	ISO 639-1 言語コード	言語ラベル	代替ラベル
アラビア語	AR	arabic	なし
チェコ語	CS	czech	なし
デンマーク語	DA	danish	なし
オランダ語	NL	dutch	なし
英語	EN	us_english	english
フィンランド語	FI	finnish	N/A
フランス語	FR	french	N/A
ドイツ語	DE	german	なし
ギリシャ語	EL	greek	なし
ヘブライ語	HE	hebrew	なし
ハンガリー語	HU	hungarian	N/A
イタリア語	IT	italian	N/A
日本語	JA	japanese	N/A
韓国語	KO	korean	N/A
リトアニア語	LT	lithuanian	なし
ノルウェー語	NO	norwegian	norweg
ポーランド語	PL	polish	なし
ポルトガル語	PT	portuguese	portugue
ロシア語	RU	russian	N/A
中国語(簡体文字)	ZH	chinese	simpchin
スペイン語	ES	spanish	なし



言語	ISO 639-1 言語コード	言語ラベル	代替ラベル
スウェーデン語	SV	swedish	なし
タイ語	TH	thai	N/A
中国語 (繁体文字)	TW	tchinese	tradchin
トルコ語	TR	turkish	N/A
ウクライナ語	UK	ukrainian	なし

## ロケール文字セット

アプリケーションとサーバのロケール定義のいずれにも文字セットがあります。アプリケーションは、データベースサーバから文字列を要求するときに文字セットを使用します。データベースサーバは、データベースの文字セットをアプリケーションの文字セットと比較し、文字セット変換が必要かどうかを判断します。データベースサーバがデータベースとクライアント間の文字セット変換を実行できない場合は、接続が失敗します。

1. SACHARSET 環境変数が設定されている場合、その値で文字セットが特定されます。  
データベースサーバが SACHARSET を使用するのには、照合の指定がない状態で新しいデータベースを作成する場合だけです。
2. 接続文字列で文字セットが指定されている場合は、その文字セットが使用されます。
3. Sybase Open Client アプリケーションは、Sybase リリースディレクトリの locales サブフォルダの locales.dat ファイルを確認します。
4. オペレーティングシステムの文字セット情報によって、次の方法でロケールを特定します。
  - Windows オペレーティングシステムの場合、最新の Windows ANSI コードページが使用されます。
  - UNIX プラットフォームでは、LC\_ALL、LC\_MESSAGES、LC\_CTYPE、LANG の各ロケール環境変数が指定された順序で確認されます。この中で最初に設定されていた環境変数の値が、文字セットの決定に使用されます。文字セットをオペレーティングシステムから特定できなかった場合は、デフォルトで iso\_1 (Windows コードページ 28591、ISO 8859-1 Latin I、ISO 8859-1 Latin-1、iso\_8859-1:1987 と呼ばれる) が使用されます。
5. それ以外のプラットフォームでは、デフォルトでコードページ cp1252 が使用されます。

## 文字セット変換について

---

SAP Sybase IQ では、同じ文字が文字セットまたはコードページ内の異なる位置に定義されている場合に、文字セット間の変換を行うことができます。

この変換を行うには、文字セット間にある程度の互換性が必要です。たとえば、EUC-JIS と cp932 の文字セット間では変換できますが、EUC-JIS と cp1252 の間ではできません。

以降では、SAP Sybase IQ がどのようにして文字セット変換を実行するかを説明します。この説明は、複数の文字セットを使用する環境にアプリケーションやデータベースを配備するユーザ向けです。

### データベースメッセージの文字変換

データベースソフトウェアから出力されるエラーなどのメッセージは、言語リソースライブラリに保持されています。ローカライズ版 SAP Sybase IQ は、このライブラリのローカライズ版を備えています。

クライアントアプリケーションのユーザには、データベースのデータとメッセージの両方が表示される可能性があります。言語ライブラリ内の文字列である一部のデータベースメッセージには、データベースからの文字を入れるプレースホルダが含まれています。たとえば、存在しないカラムを指定してクエリを実行すると、次のエラーメッセージが返されます。

```
Column column-name not found
```

*column-name* の値はデータベースから提供されます。

データベースの文字セットが言語ライブラリと異なる場合でも、このような情報をクライアントアプリケーションに統合的に提供するために、データベースサーバは、データベース照合で使用されている文字セットに合わせてメッセージの文字を自動的に変換します。

メッセージは、文字セット変換がオンになっているかオフになっているかにかかわらず、常にデータベース照合の文字セットに変換されます。

データベースサーバの文字セット変換がオン (デフォルト) になっており、かつクライアントの文字セットがデータベース照合のものと異なる場合は、さらに文字セット変換が実行されます。

#### データベースメッセージの文字変換の使用

データベースメッセージを読み取る前に確認します。

1. データベースで使用する照合が、クライアントマシンで使用されている文字セットや、SAP Sybase IQ の言語リソースライブラリで使用されている文字

セットと互換性があることを確認してください。ローカライズの異なる SAP Sybase IQ では、言語リソースライブラリも異なります。

2. 使用したいすべての文字が各文字セットに入っているか確認します。

## 接続文字列と文字セット

接続文字列の文字変換は、特殊なケースです。

接続文字列はクライアントライブラリで次のように解析され、データベースサーバが検出または起動されます。

- *keyword* = *value* コンポーネントに分解されます。これは、CommLinks パラメータを中カッコ { } で囲んでいないかぎり、文字セットの種類に関係なく実行されます。中カッコの代わりにカッコ ( ) を使用することをおすすめします。中カッコは、一部のマルチバイト文字セットで有効な後続バイト (最初のバイト以外のバイト) です。
- サーバが検索されます。サーバ名はクライアントマシンの文字セットに従って解釈されます。Windows の場合は、ANSI 文字セットが使用されます。クライアントとサーバのマシン間で文字セット変換の問題が起きなければ、拡張文字も使用できます。  
マシン間での互換性を最大限に高めるためには、サーバ名にはアルファベットの ASCII 文字 1 ~ 127 (または 33 ~ 126) およびアンダースコアのみを使い、記号を使わないようにしてください。サーバ名は 40 文字にトランケートされます。
- DatabaseName (DBN) または DatabaseFile (DBF) パラメータは、データベースサーバの文字セットで解釈されます。
- データベースが検出されると、他の接続パラメータがデータベースの文字セットに従って解釈されます。

## 文字セット変換の回避

文字セット変換にはパフォーマンスコストがかかります。文字セット変換の必要がない環境を設定できれば、このコストを考慮する必要がなく、設定の管理も簡単です。

シングルバイト文字セット (SBCS) で作業をしていて、7ビット ASCII 文字 (値 1 ~ 127) だけが使用されている場合、文字セット変換は必要ありません。データベースとクライアントのオペレーティングシステムの間でコードページが異なっている場合でも、この範囲では互換性があります。英語のインストール環境の多くは、この条件に当てはまります。文字セット変換は、デフォルトではオンになっています。

拡張文字を使用する必要がある場合は、次のとおりです。

## 国際言語サポートについて

- クライアントマシンのオペレーティングシステムのコードページが、データベース内で使用されているものと一致する場合は、データベース内のデータには文字セット変換は必要ありません。
- 自分の国の言語にローカライズされた SAP Sybase IQ が使用可能で、オペレーティングシステムでコードページを使用している場合は、データベースメッセージの文字セット変換は必要ありません。SAP Sybase IQ のメッセージ文字列で使われる文字セットは、以下のとおりです。

言語	文字セット
英語	cp1252
フランス語	cp1252
ドイツ語	cp1252
日本語	cp932 (シフト JIS)

接続文字列で **CharSet=none** を指定した場合、文字セット変換は無効になります。

## ANSI コードページと OEM コードページ

SAP Sybase IQ の照合は、ANSI または OEM として指定されたコードページに基づいています。

ほとんどの場合、ANSI コードページを使用してください。OEM を使用する場合は、ユーザのクライアントマシンの OEM コードページに合うコードページを選択します。

---

**注意：** 別の変換ドライバは使用しないでください。変換ドライバはサーバの文字セット変換と競合し、データの破壊につながる可能性があります。

---

Interactive SQL では、iAnywhere JDBC ドライバが文字セット変換を取り扱います。

### ANSI ISO\_1 照合

---

ISO\_1 は、Adaptive Server のデフォルトである ISO\_1 照合と互換性があります。相違点は次のとおりです。

- B、小文字の sharp s (¥xDF) は、SAP Sybase IQ と SQL Anywhere では小文字の **s** の後にソートされますが、Adaptive Server では **ss** の後にソートされます。
- SAP Sybase IQ と SQL Anywhere では、**Æ** と **æ** (それぞれ、¥xC6 と ¥xE6) に対応する合字は、それぞれ **A** と **a** の後にソートされます (**A**、**a**、**æ**、**Æ**)。Adaptive Server でのソート順は、**A**、**a**、**Æ**、**æ** となります。

### ANSI 1252LATIN1 照合

---

この照合には、ユーロ通貨記号やハーチェック付きの Z (Ž)、ハーチェック付きの z (ž) などいくつかの文字が含まれています。

シングルバイトの Windows オペレーティングシステムでは、この照合をおすすめします。

英語または西欧言語を使用している Windows ユーザには、この照合をおすすめします。

## ANSI ISO1LATIN1 照合

---

この照合は ISO\_1 と同じですが、A0 ~ BF の範囲にある値のソートが含まれています。

Adaptive Server Enterprise との互換性を保つために、ISO\_1 照合では 0xA0 ~ 0xBF の範囲に文字が定義されていません。ただし、この照合のベースとなった ISO Latin 1 文字セットには、この範囲にも文字が割り当てられています。ISO1LATIN1 照合では、この範囲の文字が反映されます。

Adaptive Server Enterprise と互換性を保つ必要がない場合、通常は、ISO\_1 ではなく ISO1LATIN1 を使用することをおすすめします。

英語または西欧言語を使用している UNIX ユーザで、デフォルト照合である ISO\_BINENG の最適なパフォーマンスが多少損なわれてもよい場合は、ISO1LATIN1 をおすすめします。

## ANSI ISO9LATIN1 照合

---

この照合は ISO1LATIN1 と同じですが、1252LATIN1 照合に含まれる、ユーロ通貨記号などの新しい文字を含んでいます。

お使いのコンピュータで ISO Latin 9 文字セットを使用していて、ISO\_BINENG の最適なパフォーマンスが多少損なわれてもよい場合は、ISO9LATIN1 を使用してください。

## 照合の表示

現在使用されている照合を取得します。  
コマンドプロンプトで次のように入力します。

```
SELECT * FROM SYS.SYSCOLLATION
```

照合の表示



# マルチバイト照合

SAP Sybase IQ には、いくつかのマルチバイト文字セットを使用した照合が用意されています。

SAP Sybase IQ は可変幅文字セットをサポートしています。可変幅の文字セットには、1バイトで表される文字と、複数バイト (最大4バイト) で表される文字が混在しています。文字の最初のバイトは、その文字に使用するバイト数と、その文字の種類 (スペース、数字、またはアルファベット) を示します。

UTF8 照合については、UTF-8 文字は 1～4 バイトで表されます。他のマルチバイト照合の場合、1 または 2 バイトが使用されます。用意されているすべてのマルチバイト照合についても、2 バイト以上で構成される文字は「アルファベット」とみなされ、二重引用符なしで識別子に使用できます。

SAP Sybase IQ は、UTF-16 や UTF-32 など、16 ビットおよび 32 ビットの文字セットをサポートしません。

Embedded SQL 以外のすべてのクライアントライブラリは Unicode に対応しており、UTF-16 エンコーディングが使用できます。クライアントとサーバの間で変換が実行されます。

## 日本語のサポート

---

日本語の Windows アプリケーションには照合 932JPN を、日本語の UNIX アプリケーションには EUC\_JAPAN を使用してください。

照合 932JPN は、古い照合の SJIS や SJIS2 にはロードできない 32 ビットマルチバイト文字のロードをサポートしています。SJIS は引き続きサポートされていますが (32 ビットのマルチバイト文字を除く)、SJIS2 は現在ではサポートされていません。

## タイ語のサポート

---

SAP Sybase IQ では、**CP874toUTF8** ユーティリティを使用して、CP874 形式のデータファイルを UTF8 に変換します。これは、タイ語のために SAP Sybase IQ でサポートされる照合です。

このユーティリティを使用して、データを UTF8 に変換せずに CP874 文字セットのデータをロードすることもできます。

## マルチバイト照合

**SORTKEY()** 関数は、UTF8 形式のタイ語の文字セットであるソート順 **thaidict** (タイ語辞書) で値を返します。次の各文はどれも同じ結果になります。

```
SELECT c1, SORTKEY(c1) from T1 where rid=3
SELECT c1, SORTKEY(c1, 'thaidict') from T1 where rid=3
SELECT
'¥340¥270¥201¥340¥271¥207', SORTKEY('¥340¥279¥201¥340¥271¥207') from
T1 where rid=3
```

## ロケール情報

ロケール情報は、PROPERTY、DB\_PROPERTY、CONNECTION\_PROPERTYなどの関数を使用して確認できます。次の表は、これらの関数を使用して、クライアント接続、データベース、データベースサーバのロケール情報を返す方法を示します。

システム関数とパラメータ	戻り値
SELECT PROPERTY( 'CharSet' );	データベースサーバの文字セット。通常は、データベースサーバを実行しているコンピュータの文字セット
SELECT PROPERTY( 'DefaultCollation' );	データベースサーバがデータベース作成時に使用するデフォルトの CHAR 照合
SELECT PROPERTY( 'DefaultNcharCollation' );	データベースサーバがデータベース作成時に使用するデフォルトの NCHAR 照合
SELECT PROPERTY( 'Language' );	データベースサーバのロケール言語
SELECT DB_PROPERTY( 'CharSet' );	CHAR データをデータベースに格納するときに使用する文字セット
SELECT DB_PROPERTY( 'NcharCharSet' );	NCHAR データをデータベースに格納するときに使用する文字セット
SELECT DB_PROPERTY( 'MultiByteCharSet' );	CHAR データでマルチバイト文字セットを使用するかどうか (On=使用、Off=使用しない)
SELECT DB_PROPERTY( 'Language' );	データベースの CHAR 照合でサポートしている言語を表す 2 文字のコードをカンマで区切ったリスト
SELECT DB_PROPERTY( 'Collation' );	データベースサーバが使用する CHAR 照合名
SELECT DB_PROPERTY( 'NcharCollation' );	データベースサーバが使用する NCHAR 照合名
SELECT CONNECTION_PROPERTY( 'CharSet' );	クライアントの CHAR データ文字セット
SELECT CONNECTION_PROPERTY( 'NcharCharSet' );	接続に使用する NCHAR データの文字セット

## ロケール情報

システム関数とパラメータ	戻り値
SELECT CONNECTION_PROPERTY( 'Language' );	クライアントが接続に使用する言語

## 特定のロケールをサポートする照合の選択

文字セットの問題に適切に解決するには、ロケールを最適にサポートする照合を使用します。

1. その環境で使用する各コンピューティングプラットフォームについて、デフォルトロケールを調べます。デフォルトロケールは、各コンピュータの文字セットと言語です。Windows では、文字セットは ANSI コードページです。
2. ご自分の環境にそのロケール設定が適しているかどうか確認します。
3. デフォルトの設定が不適切な場合は、データに合った文字セット、言語、データベース照合を選択し、文字セット変換が行われないようにしてください。
4. 環境内の各コンピュータのロケールを、決定した値に設定します。
5. デフォルトの照合を使用してデータベースを作成するか、デフォルトの照合が適切でない場合は、名前付きの照合を使用します。
6. データベースで使用する照合を選択します。
  - a) データに適した文字セットとソート順を使用する照合を選択します。複数の照合 (OEM や ANSI) がこの条件を満たす場合があります。
  - b) 文字セット変換を行わずにすむ照合を選択してください。文字セット変換はパフォーマンスコストがかかるうえ、システム設定も複雑になります。クライアントマシンのオペレーティングシステムで使用している文字セットに合った照合順をデータベースで使用すると、文字セット変換が不要になります。

特定のロケールをサポートする照合の選択

# 文字セットと照合順の選択によるパフォーマンスの最大化

文字セットと照合順がバイナリの場合の方が、バイナリでない場合より文字データのパフォーマンスが向上します。

パフォーマンスを最大にするには、次のようにデフォルトのオプション設定でデータベースを作成します。これにより、文字セットと照合順がバイナリになります。

```
CREATE DATABASE dbname  
COLLATION 'ISO_BINENG'  
CASE RESPECT
```

これらのオプションを上記以外の設定にすると、文字セットも照合順もバイナリになりません。

この設定の短所は、大文字が常に小文字の前に置かれることです。たとえば、"apple" よりも "BANANA" の方が前になります。より自然なソート順にしながら、大文字と小文字を区別するデータベースが必要な場合、パフォーマンスが多少損なわれてもよければ、デフォルトの ISO\_BINENG ではなく ISO\_1 の照合を使用してください。

---

**注意：**新しく作成されたデータベースでは、データベースの設定にかかわらず、パスワードの大文字と小文字が区別されます。既存のデータベースを再構築する場合、パスワードの大文字と小文字の区別は元のデータベースの設定に準じます。

## 文字セットと照合順の選択によるパフォーマンスの最大化



## ロケールの設定

オペレーティングシステムのデフォルトロケールを使用することも、コンピュータ上の SAP Sybase IQ コンポーネントで使用するロケールを明示的に指定することもできます。

1. IQ エンジンの文字セットを表示します。

```
select * from sa_eng_properties() where proptime like '%CharSet%'
```

2. IQLANG 環境変数と IQCHARSET 環境変数のいずれかまたは両方を設定します。

```
IQCHARSET=charset;IQLANG=language_code
```

*charset* は有効な文字セットラベルで、*language\_code* は言語ラベルの値のリストにある言語コードです。

SAP Sybase IQ は、**INSERT...LOCATION** 文の実行時にローカライゼーション情報をロードします。

## 言語ラベル値のリスト

有効な言語ラベルの値と対応する ISO 639 言語コードは次のとおりです。IQ\_LANG 環境変数に 2 文字の ISO\_639 言語コードを設定してください。

言語	ISO_639 言語コード	言語ラベル	代替ラベル
アラビア語	AR	arabic	該当なし
チェコ語	CS	czech	該当なし
デンマーク語	DA	danish	該当なし
オランダ語	NL	dutch	該当なし
英語	EN	us_english	english
フィンランド語	FI	finnish	該当なし
フランス語	FR	french	該当なし
ドイツ語	DE	german	該当なし
ギリシャ語	EL	greek	該当なし
ヘブライ語	HE	hebrew	該当なし

## ロケールの設定

言語	ISO_639 言語コード	言語ラベル	代替ラベル
ハンガリー語	HU	hungarian	該当なし
イタリア語	IT	italian	該当なし
日本語	JA	japanese	該当なし
韓国語	KO	korean	該当なし
リトアニア語	LT	lithuanian	該当なし
ノルウェー語	NO	norwegian	norweg
ポーランド語	PL	polish	該当なし
ポルトガル語	PT	portuguese	portugue
ロシア語	RU	russian	該当なし
中国語(簡体字)	ZH	chinese	simpchin
スペイン語	ES	spanish	該当なし
スウェーデン語	SV	swedish	該当なし
タイ語	TH	thai	該当なし
中国語(繁体字)	TW	tchinese	tradchin
トルコ語	TR	turkish	該当なし
ウクライナ語	UK	ukrainian	該当なし

## INSERT...LOCATION 文のロケールの設定

データベースが使用するロケールがプラットフォームのデフォルトと異なる場合は、ローカルクライアントの環境変数を設定する必要があります。言語、照合順、文字セット、日付/時刻フォーマットの正しい情報を SAP Sybase IQ がロードできるようにするためです。

SAP Sybase IQ は、ロケール名を決定するときに、まず LC\_ALL 環境変数の値をチェックします。LC\_ALL が設定されていない場合、SAP Sybase IQ は LANG 環境変数の値を使用します。どちらの環境変数も設定されていない場合、SAP Sybase IQ はロケールファイルにある「デフォルト」のエントリを使用します。

1. テキストエディタで \$SYBASE/locales/locales.dat ファイルを開きます。

次に例を示します。

```
locale = default, us_english, roman8  
locale = C, us_english, roman8  
locale = american, us_english, roman8  
locale = english.iso88591, us_english, iso_1
```

2. LC\_ALL または LANG 環境変数を正しい値に設定します。手順 1 のプラットフォームで、データベースの照合が iso\_1、使用する言語が英語の場合は、環境変数 LC\_ALL または LANG の値を "american.iso88591" に設定します。設定しない場合、SAP Sybase IQ はロケール名 "default" を使用します。その照合は "roman8" です。

たとえば、sh シェルまたは ksh シェルの場合、次のように入力します。

```
LC_ALL= american.iso88591;export LC_ALL
```

csh シェルまたは tsch シェルの場合、次のように入力します。

```
setenv LC_ALL american.iso88591
```



## データベースサーバで文字セット変換を無効にする

データベースサーバのコマンドラインを使用すれば、文字セット変換のオンとオフを明示的に切り替えることができます。

SAP Sybase IQ では、文字セット変換がデフォルトで有効になっており、クライアントとサーバのロケールが異なる場合に実行されます。

文字セット変換を無効にするには、接続文字列で `CharSet=none` を入力します。

データベースサーバで文字セット変換を無効にする

# デフォルトの照合を使用したデータベースの作成

デフォルトの操作で SAP Sybase IQ データベースを作成すると、そのデータベースの照合は ISO\_BINENG になります。

ISO\_BINENG 照合を使用すると、IQ データベースで最適なパフォーマンスが発揮されますが、ソート順は必ずしも適切ではありません。

これは SQL Anywhere と異なり、データベースを作成するオペレーティングシステムの文字セットに応じて、新しいデータベースのデフォルトの照合が類推されるわけではありません。

このデフォルトの方法でシステムを設定できない場合は、データベースで使用する照合を決定し、データベースシステム間で整合性を保ったままデータを交換できるように文字セット変換を使用するかどうかを判断してください。

## デフォルトの照合を使用したデータベースの作成



# データベースの照合の変更

選択した照合へのアンロードおよび再ロードを実行することで、ある照合から別の照合へ、データベースの照合を変更できます。照合はデータベースの作成時に選択されるものであり、データベースを再構築せずに変更することはできません。

## 前提条件

SELECT ANY TABLE および SERVER OPERATOR システム権限が必要です。

デフォルトでは、UNLOAD 文を実行するには SELECT ANY TABLE システム権限が必要です。-gl データベースサーバオプションを使用すると、必要な権限を変更できます。

## 手順

1. データベースを起動します。

```
START databasefile
```

2. 既存のデータベースの文字セットを確認します。

```
SELECT DB_PROPERTY( 'CharSet' );
```

SAP Sybase IQ の初期バージョンでは、このプロパティがない場合があります。文字セットは照合名で判断することもできます。たとえば、照合 1252LATIN1 はコードページ 1252 を使用しています。

既存のデータベースの文字セットは、オペレーティングシステムおよびクライアントの文字セットと同じか、互換性があります。そうでない場合は、データベースの再構築が必要になりますが、再構築のプロセスには細心の注意が必要です。

特に、使用しているデータベースの照合が 850LATIN1 であり、使用している SAP Sybase IQ が初期バージョンであるため文字セット変換がサポートされていないか (バージョン 5 以前) デフォルトで無効になっており (バージョン 6 と 7)、クライアントアプリケーションが標準的な Windows アプリケーションの場合、データベースにコードページ 1252 の文字データが含まれている可能性があります (通常はコードページ 850 に含まれる解釈データです)。

3. 既存のデータベースに格納されているデータの文字セットを確認します。

```
UNLOAD TABLE mytable TO 'mytable-data-in-utf8.dat' ENCODING 'UTF-8';
```

結果をテキストエディタで表示します。

アクセント記号付きデータが正しい場合、データベースに含まれている文字データは Windows ANSI コードページに対応しており、英語と西ヨーロッパ言

## データベースの照合の変更

語の場合はコードページ 1252 です。データが DOS ベースのエディタで正常に表示される場合、文字データは Windows OEM コードページに対応しており、通常は 437 または 850 です。

### 4. データベースをアンロードします。

データの文字セットがデータベースの文字セットと互換性がない場合、文字セット変換なしでデータをアンロードすることが重要です。使用されている SAP Sybase IQ によっては、`dbunload` の内部アンロード機能を使用したり、`UNLOAD TABLE` 文を使用してデータを手動でアンロードしたりできます。

### 5. 新しいデータベースを作成し、使用する照合と文字セットを指定し、DBA ユーザ ID とパスワードを DBA および SQL に設定します。

```
iqinit -dba DBA,sql -z 1252LATIN1 c:¥newdatabase.db
```

### 6. 古いデータベースサーバを停止して、新しいデータベースサーバを起動します。

```
iqsrv16 -n new-server c:¥newdatabase.db
```

### 7. データを新しいデータベースにロードします。

アンロードしたデータとスキーマ (`reload.sql`) が再ロードに使用するコンピュータの文字セットに対応している場合、`dbunload` の外部再ロードオプションを使用できます。データは、データベースの正しい文字セットに自動的に変換されます。次に例を示します。

```
dbunload -ix c:¥databasefile
```

データのエンコードがデータベースの文字セットと一致してなく、データのロードに `LOAD TABLE` 文 (内部再ロード) を使用している場合は、`ENCODING` 句を使用する必要があります。データベースサーバは、`LOAD TABLE` 文を使用してロードされたデータに対し、デフォルトでは文字セット変換を実行しません。

データのエンコードが作業に使用しているコンピュータのコードページと一致せず、ロードに `INPUT` 文 (外部再ロード) を使用している場合は、`ENCODING` 句を使用する必要があります。そうしないと、データベースサーバはデータがコンピュータのネイティブ文字セットであると想定します。

新しい照合でデータベースが再作成されます。

# 索引

## A

ANSI コードページ  
    選択 25  
ASCII  
    文字セット 4

## C

CREATE DATABASE 文  
    照合 1

## I

ICU ライブラリ 1  
IQCHARSET 環境変数 37  
IQLANG 37  
IQLANG 環境変数 37  
ISO\_1 照合  
    説明 25  
ISO\_BINENG  
    デフォルトの照合 35, 43  
ISO\_BINENG 照合 43

## L

LANG 環境変数 38  
language\_code 37  
LC\_ALL 環境変数 38

## O

OEM コードページ  
    選択 25

## S

SORT\_COLLATION オプション 1  
SORTKEY 関数 1

## U

Unicode  
    文字セット 1, 29

Unilib ライブラリ 1

## い

移行  
    照合 1  
    ソートキー 1  
    廃止された照合 1

## え

エラーメッセージ  
    文字セット変換 22

## お

大文字と小文字の区別  
    パスワード 35

## か

カスタム照合 1  
環境変数  
    LANG 38  
    LC\_ALL 38

## こ

コードページ  
    サポート対象 4  
    説明 4

## し

シングルバイト文字セット  
    説明 4

## そ

ソート  
    照合 8

## 索引

ソートキー  
    移行 1  
ソート順  
    照合 1

## て

データベース  
    文字セット 22

## は

パスワード  
    大文字と小文字の区別 35

## ま

マルチバイト文字セット  
    使用 29

## ゆ

ユーティリティ  
    CP874toUTF8 1  
ユーロ記号  
    1252LATIN1 照合 25, 26

## ろ

ロケール  
    INSERT...LOCATION 38  
    設定 37  
    説明 1  
    文字セット 22