

補足資料 -2000-5-21 HiBase メタインデックス機能

2000.5.21 西林瑞夫 / ホロン株式会社

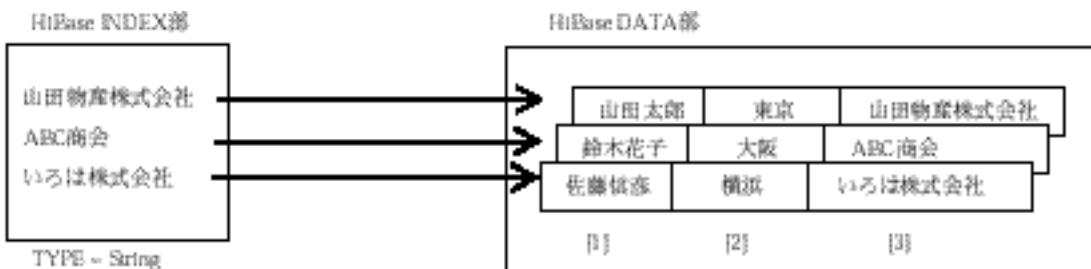
本文は、HiBase メタインデックス機能の目的と具体的な利用の方法を解説するものです。

HiBase メタインデックス機能は、HiBase-415 Rev-000521 以降の HiBaseデータベース管理システムで有効です。

目的：

HiBase メタインデックス機能は、HiBaseデータベースの検索を高速化するための手法を提供するものです。

HiBaseデータベースは、下図のようにデータ部とインデックス部から構成されます。



このうち、検索に利用されるのはインデックス部で、これはデータ部にある特定の項目のインバーテッドインデックスとなっており、B-Tree 構造で作られています。

B-Tree 構造は、「予めソートした状態で値を格納する」という手法で、参照/更新の両面でバランスのとれた汎用的な技術といえますが、例えば後方一致検索のような特殊な検索においては役に立ちません。

メタインデックス機能は、この B-Tree 構造の役に立たない分野を補完するもので、基本的にはデータ部の値を検索に都合の良い値に変換して（値をメタ空間に射影）インデックスを作りだします。

このメタインデックスの手法は、汎用的な技術といえますが、今回（HiBase415-000521）のリリースは、「後方一致検索」「アルファベットのケース無視検索」「日本語のひらがな/カタカナ無視検索」の高速化オプションを提供します。

方法：

例えば、HiBaseデータベースファイルのスキーマ定義を以下のようにします。

```
[ITEM: 1] TYPE = Number LENGTH = 6
[ITEM: 2] TYPE = String LENGTH = 6
[ITEM: 3] TYPE = String LENGTH = 6
[ITEM: 4] TYPE = String LENGTH = 6
[ITEM: 5] TYPE = String LENGTH = 6

[KEY: 1] TYPE = "Number, Unique" BIND = [1]
[KEY: 2] TYPE = "String, Reverse" BIND = [2]
[KEY: 3] TYPE = "String, IgnoreCase" BIND = [3]
[KEY: 4] TYPE = "String, IgnoreJCase" BIND = [4]
[KEY: 5] TYPE = "String, Unique, Reverse, IgnoreCase, IgnoreJCase" BIND = [5]
```

・解説

[KEY:1]は、インデックスを数（数字の値）として作りだします。また、「値の重複を排除する」というオプションが指定されています。

「値の重複を排除する」は、今回のリリースではなく、以前からあったものです。

[KEY:2]は、インデックスを文字列として作りだし、「後方一致検索の高速化オプション」が指定されています。この指定によって、「[6] = ?"株式会社"」のような後方一致が早くなります。
[KEY:3]は、インデックスを文字列として作りだし、「アルファベットのケース無視検索の高速化オプション」が指定されています。この指定によって、「[6] = "abc", "ABC"」のような英字のケースを意識する必要がなくなります。

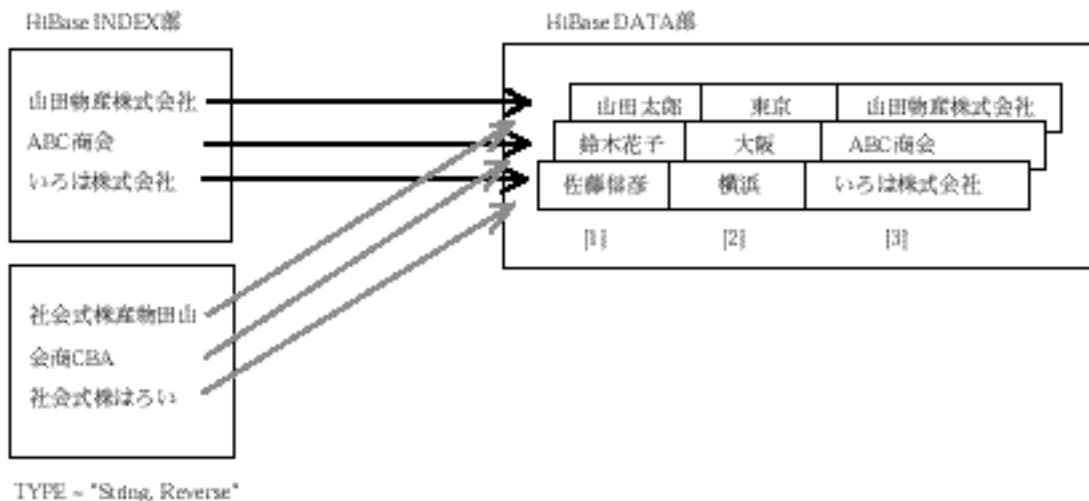
[KEY:4]は、インデックスを文字列として作りだし、「日本語のひらがな/カタカナ無視検索の高速化オプション」が指定されています。この指定によって、「[6] = "やまだ", "ヤマダ"」のようなカタカナ/ひらがなの字種を意識する必要がなくなります。

[KEY:5]は、全てのオプションを指定する場合の例です。

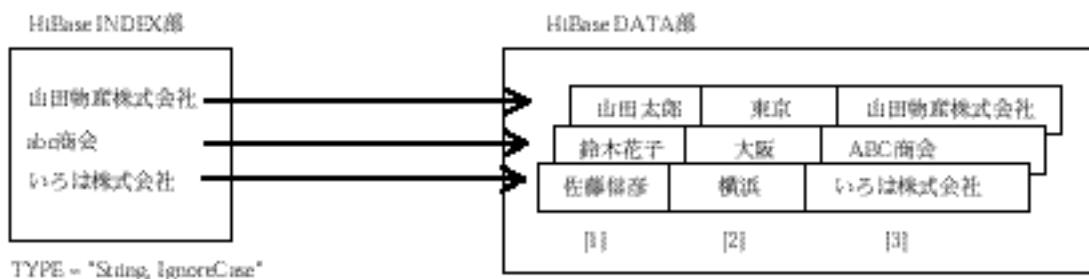
・参考

上記の、「"String,Reverse"」、「"String,IgnoreCase"」、「"String,IgnoreJCase"」でつくられた影響を以下の図に示します。

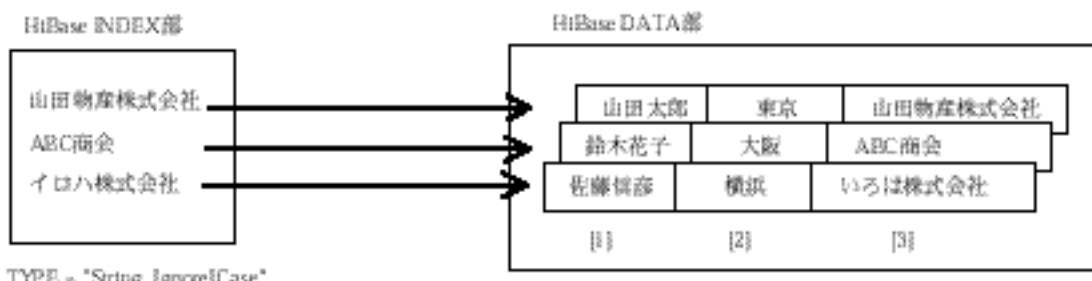
後方一致検索の高速化オプション



アルファベットケース無視の高速化オプション



カタカナ / ひらがな無視の高速化オプション



その他：

1. HiBase メタインデックス機能は、HiBase 415 Rev-000521 以降に利用できる機能です。
2. HiBase メタインデックス機能は、Windows 95/98/NT/2000、Macintosh、UNIXの全ての環境で有効です。またデータベースと環境設定ファイルも全ての環境でバイナリ互換です（例えば、Macintoshで作ったデータベースと環境設定ファイルをそのまま UNIX にコピーして運用することができます）