<https://mp.weixin.qq.com/s/X_OV4aUeJiFN9DIWMNpICA>

# 为什么不建议在MySQL中使用 utf8 ？

MySQL 字符编码集中有两套 UTF-8 编码实现：utf8 和 utf8mb4。

如果使用 utf8 的话，存储 emoji 符号和一些比较复杂的汉字、繁体字就会出错。

## 何为字符集？

字符是各种文字和符号的统称，包括各个国家文字、标点符号、表情、数字等等。**字符集** 就是一系列字符的集合。字符集的种类较多，每个字符集可以表示的字符范围通常不同，就比如说有些字符集是无法表示汉字的。

**计算机只能存储二进制的数据，那英文、汉字、表情等字符应该如何存储呢？**

我们要将这些字符和二级制的数据一一对应起来，比如说字符“a”对应“01100001”，反之，“01100001”对应 “a”。我们将字符对应二进制数据的过程称为"**字符编码**"，反之，二进制数据解析成字符的过程称为“**字符解码**”。

Unicode 字符集中包含了世界上几乎所有已知的字符。不过，Unicode 字符集并没有规定如何存储这些字符（也就是如何使用二级制数据表示这些字符）。

然后，就有了 **UTF-8**（**8**-bit **U**nicode **T**ransformation **F**ormat）。类似的还有 UTF-16、 UTF-32。

UTF-8 使用 1 到 4 个字节为每个字符编码， UTF-16 使用 2 或 4 个字节为每个字符编码，UTF-32 固定位 4 个字节为每个字符编码。



