

后盾网 人人做后盾

www.houdunwang.com

字符集与校对规则

后盾网 2011-2016 v5.0

字符集

- 字符集(Character set)是多个字符的集合，字符集种类较多，每个字符集包含的字符个数不同，常见字符集名称：ASCII字符集、GB2312字符集、BIG5字符集、GB18030字符集、Unicode字符集等
- 计算机要准确的处理各种字符集文字，需要进行字符编码，以便计算机能够识别和存储各种文字。中文文字数目大，而且还分为简体中文和繁体中文两种不同书写规则的文字，而计算机最初是按英语单字节字符设计的，因此，对中文字符进行编码，是中文信息交流的技术基础。

字符集

- GBK编码是中国大陆制订的、等同于UCS的新的中文编码扩展国家标准。该编码标准兼容GB2312，共收录汉字21003个、符号883个，并提供 1894个造字码位，简、繁体字融于一库。P-Windows3.2和苹果OS以GB2312为基本汉字编码， Windows 95/98则以GBK为基本汉字编码。
- GBK采用双字节表示

GBK编码

- UTF-8是UNICODE的一种变长字符编码又称万国码，UTF-8用1到3个字节编码UNICODE字符。在网页上可以同一页面显示中文简体繁体及其它语言(如日文，韩文)
- 如果要制作多语言版本的网站，请选用UTF-8字符集

UTF8

show character set

- 查看服务器支持哪些字符集

show collation

- 字符校对排序规则，如果没有指定相应字符集校对规则，将使用默认方式进行校对与排序

show variables like "%character%";

- 查看系统字符集设置变量

set names utf8

- 设置客户端和服务端连接使用的字符集

字符集

show collation

- 查看MySQL支持的校对规则
 - 当没有选择具体校对规则时，为YES的为默认较对规则
 - 可以通过show character set 查看到默认校对规则
 - 校对规则可以是区分大小写的或不区分的或者是二进制的
-
- `create table a(name varchar(30) character set utf8 collate utf8_bin,name2 varchar(30) character set utf8 collate utf8_general_ci);`
 - `insert into a(name,name2)values("a","a"),("b","b"),("A","A"),("B","B");`
 - `select * from a where name='a';`//只能找到一个
 - `select * from a order by name;`//name区分大小写校对的排序

校对规则

- 如果指定了字符集character set而没有指定collate校对规则，校对规则会使用字符集的默认校对规则
- 如果指定了校对规则而没有指定字符集，将使用当前校对规则的字符集
- 如果没有指定校对规则也没有指定字符集，将使用表的校对规则与字符集
- 如果表没有指定字符集与校对规则，将使用数据库的字符集与校对规则
- 如果数据库没有设置校对规则与字符集，将使用服务器的默认校对规则与字符集规则

字符集及校验设置规则
