



MySQL
Kenneth Power

Simplify.



All trademarks used herein are the sole property of their respective owners.

Simplify.

Topics

- »» Character Sets & Collations
- »» Resource Management

Simplify.

Character Sets & Collations

Simplify.

Questions

- »» What causes this?
- »» Is it preventable?
- »» How can it be fixed?

Basics

- »» What is a character set?
- »» What is a collation?

Character set

A set of symbols and encodings

Symbol: what we recognize
example: a, B, t, z

Encoding: how the symbols appear to the
computer

Collation

A set of rules for comparing symbols

ci: case insensitive

cs: case sensitive

bin: binary value comparison



What Does Your Server Have?

Simplify.

What does your server have?

```
mysql> SHOW CHARACTER SET;
```

Charset	Description	Default collation	Maxlen
big5	Big5 Traditional Chinese	big5_chinese_ci	2
dec8	DEC West European	dec8_swedish_ci	1
cp850	DOS West European	cp850_general_ci	1
hp8	HP West European	hp8_english_ci	1
koi8r	KOI8-R Relcom Russian	koi8r_general_ci	1
latin1	cp1252 West European	latin1_swedish_ci	1
latin2	ISO 8859-2 Central European	latin2_general_ci	1
swe7	7bit Swedish	swe7_swedish_ci	1
ascii	US ASCII	ascii_general_ci	1
ujis	EUC-JP Japanese	ujis_japanese_ci	3
sjis	Shift-JIS Japanese	sjis_japanese_ci	2
hebrew	ISO 8859-8 Hebrew	hebrew_general_ci	1
tis620	TIS620 Thai	tis620_thai_ci	1
euckr	EUC-KR Korean	euckr_korean_ci	2
koi8u	KOI8-U Ukrainian	koi8u_general_ci	1
gb2312	GB2312 Simplified Chinese	gb2312_chinese_ci	2
greek	ISO 8859-7 Greek	greek_general_ci	1
cp1250	Windows Central European	cp1250_general_ci	1
gbk	GBK Simplified Chinese	gbk_chinese_ci	2

What does your server have?

```
mysql> SHOW COLLATION LIKE 'utf8%';
```

Collation	Charset	Id	Default	Compiled	Sortlen
utf8_general_ci	utf8	33	Yes	Yes	1
utf8_bin	utf8	83		Yes	1
utf8_unicode_ci	utf8	192		Yes	8
utf8_icelandic_ci	utf8	193		Yes	8
utf8_latvian_ci	utf8	194		Yes	8
utf8_romanian_ci	utf8	195		Yes	8
utf8_slovenian_ci	utf8	196		Yes	8
utf8_polish_ci	utf8	197		Yes	8
utf8_estonian_ci	utf8	198		Yes	8
utf8_spanish_ci	utf8	199		Yes	8
utf8_swedish_ci	utf8	200		Yes	8
utf8_turkish_ci	utf8	201		Yes	8
utf8_czech_ci	utf8	202		Yes	8
utf8_danish_ci	utf8	203		Yes	8
utf8_lithuanian_ci	utf8	204		Yes	8
utf8_slovak_ci	utf8	205		Yes	8
utf8_spanish2_ci	utf8	206		Yes	8
utf8_roman_ci	utf8	207		Yes	8

For every character there is a default

Simplify.

What Determines the Default

1. Server
2. Database
3. Table
4. Column
5. Connections

Server

```
--default-character-set=charset  
--default-collation=collation
```

Database

```
mysql> CREATE DATABASE db_name CHARACTER SET charset  
      COLLATE collation;
```

```
mysql> ALTER DATABASE db_name CHARACTER SET charset  
      COLLATE collation;
```

Checks:
 Server

Table

```
mysql> CREATE TABLE tbl_name CHARACTER SET charset  
      COLLATE collation;
```

```
mysql> ALTER TABLE tbl_name CHARACTER SET charset  
      COLLATE collation;
```

Checks:
Database
Server

Column

```
mysql> col_name {CHAR | VARCHAR | TEXT} CHARACTER SET  
charset COLLATE collation;
```

```
mysql> ALTER TABLE ...;
```

Checks:
Table
Database
Server

Connection

`character_set_client`

Determines character set the statement is in when it leaves the client

`character_set_connection`

Determines character set the statement is translated to by server

Connection

`collation_connection`

Determines collation used when performing comparisons

`character_set_results`

Determines character set the server should translate to before sending result sets or error messages to the client

Connection

How does one set these variables?

```
mysql> SET NAMES 'charset_name';
```

```
mysql> SET character_set_client = x;
```

```
mysql> SET character_set_results = NULL;
```

Application, language specific

When Nothing is Defined

latin1

Simplify.

Summary

Character set

Collation

Simplify.

What Causes This?

Upgrade from MySQL 4.0

Before 4.1.0, a character set included symbols,
encodings and collations

Poorly designed End user applications

Misuse of tools

How can it be fixed?

What is causing the issue?

How many character sets are involved?

Simplify.

Root Cause

```
| character_set_results | utf8 |
```

```
mysql> set character_set_results = latin1;
```

```
mysql> select * from t;
```

```
+-----+
```

```
| blog_content |
```

```
+-----+
```

```
| Нас раннего |
```

```
+-----+
```

```
1 row in set (0.00 sec)
```

Investigation Needed

Simplify.

How do we fix it?

Simplify.

Single Character Set

```
CREATE TABLE t(  
    ...  
    blog_content TEXT;  
    ...  
);
```

```
ALTER TABLE t MODIFY blog_content BINARY;  
ALTER TABLE t MODIFY blog_content TEXT CHARACTER SET utf8;
```

Multiple Character Sets

1. Identify the character sets
2. Extract into separate files

```
mysql> set character_set_results=NULL;  
mysql> SELECT ... INTO OUTFILE 'file-charset.sql';  
# iconv -f 'charset' -t UTF-8 file-charset.sql > file-utf8.sql
```

3. Replace 'latin1' with 'utf8'

UTF-8 Editors

- »» Windows: Notepad
- »» Mac OS X: Text Edit
- »» Linux: Vi(m); Emacs

Multiple Character Sets

4. Replace data with converted set

Recommendations

Server default to utf8:
default-character-set=utf8
default-collation=utf8_bin

Use up-to-date, properly designed applications

Warnings

init-connect: SET NAMES binary;

Read, Read, Read

Experiment

Simplify.



Resource Management

Simplify.

Resources

» CPU

» RAM

» Disk

Simplify.

Defaults

Simplify.

Some Defaults

Variable	Default
table_cache	64
max_connections	100
key_buffer_size	8 MB

Resource tuning can have
cascading effects

max_connections

Default: 100

500

table_cache : 64

Recommended: *max_connections* x **N**

Open File Descriptors (fd)

MyISAM: 2 fd / table / connection

32,000 :: 10,000 -> 5,000 table/connections

Analysis

Simplify.

Tools

MySQL Forge <http://forge.mysql.com/>

MySQL Manual

<http://dev.mysql.com/doc/refman/5.0/en/index.html>

skip-innodb

Simplify.



Questions?

Simplify.