latin1에서_utf8mb4로_변환하기 (Migration)

- description : Database Encoding Migration
- author : 도봉산핵주먹
- email : hylee@repia.com
- lastupdate : 2020-04-22

변환한 이유

> 에러내용 (내부 프로젝트 코** 본문태그 전체를 수집해서 DB에 Insert 할때 생김)

- ERROR 1366 (HY000) : incorrect string value : '\xED\x95\x9C\xEC\x9A\xB0...' for column 'ARTCL_HTML' at row 1

> 분석

- 소스 중 이모티콘이 포함되어있어 생기는 에러
- 이모티콘 4byte라 이보다 작은 CharacterSet 인 문자 집합으로 인코딩 되어있으면 깨져서 에러가 난다.

> 참고

- mariaDB utf8은 3byte 이므로 이모티콘을 포함하려면 utf8mb4로 변환해야한다.
- 문자 집합(Character Set)과 인코딩(Encoding)

작업 순서

```
1. 현재 DB Character Set 을 확인한다. (utf8mb4인지 확인)
2. DB를 dump 한다.
3. mariaDB 설정파일에 인코딩 설정을 변경(, 추가) 한다.
4. dump 파일에 DDL 설정이 utf8mb4인지 확인하고 아니면 utf8mb4로 수정한다.
5. DB에 dump 파일로 data를 복구한다.
6. mariaDB 재시작 후 DB 인코딩 확인한다.
```

1. Character Set 확인

> Table Caracter Set 확인

```
    SELECT CCSA.character_set_name FROM information_schema.`TABLES` T,
        information_schema.`COLLATION_CHARACTER_SET_APPLICABILITY` CCSA
    WHERE
        CCSA.collation_name = T.table_collation
    AND
        T.table_schema = "${Database name}"
    AND
```

```
        T.table_name = "${Table name}";
```

- MySQL database 와 table 의 character set 확인하는 법

2. Database dump

```
[root@mysql ~]# mysqldump -uroot -p ${Database Name} -r ${Backup File Name}.sql
Password:
```

2-1. Dump File Encoding 확인

- file -bi를 하면 그 파일의 인코딩이 확인된다.
- Dump파일이 utf8 이여야지 DB를 utf9mb4로 변환했을때 깨짐없이 들어간다.

```
 [root@mysql ~]# file -bi ${Backup File Name}.sql
```

3. mariaDB 설정파일에 인코딩 설정 추가

> 위치

* /etc/my.cnf

> 추가 코드

```
#
# This group is read both both by the client and the server
# use it for options that affect everything
#
[client-server]

#
# include all files from the config directory
#
[mysqld]
default_storage_engine=innodb
innodb_buffer_pool_size=445M
innodb_log_file_size=50M
max_connections = 10000
connect_timeout = 600
wait_timeout = 6000
max_allowed_packet = 2G
thread_cache_size = 128
sort_buffer_size = 4M
bulk_insert_buffer_size = 16M
tmp_table_size  = 32M
max_heap_table_size = 32M
default_storage_engine = InnoDB
```

```
innodb_buffer_pool_size = 4G
innodb_log_buffer_size  = 8M
innodb_file_per_table   = 1
innodb_file_format      = barracuda
innodb_open_files       = 400
innodb_io_capacity      = 400
innodb_flush_method     = O_DIRECT

character-set-server=utf8mb4
collation-server=utf8mb4_general_ci

[client]
port=${Port}
default-character-set = utf8mb4

[mysqldump]
default-character-set   = utf8mb4

[mysql]
default-character-set   = utf8mb4

!includedir /etc/my.cnf.d
```

4. dump 파일 DDL

- Dump 파일을 열어보면 DDL 코드가 있다.

```
DROP TABLE IF EXISTS `${Table name}`;
/*!40101 SET @saved_cs_client     = @@character_set_client */;
/*!40101 SET character_set_client = utf8 */;
CREATE TABLE `${Table name}` (
  `${column}` ${TYPE} NOT NULL,
  `${column}` ${TYPE} DEFAULT NULL,
  `${column}` ${TYPE} DEFAULT NULL,
  `${column}` ${TYPE} DEFAULT NULL,
  `${column}` ${TYPE} DEFAULT NULL,
  PRIMARY KEY (${COLUMN}`)
) ENGINE=InnoDB DEFAULT CHARSET=utf8;
```

- 위 DDL을 보면 마지막에 `CHARSET=utf8;`가 있는데 이것을 `CHARSET=utf8mb4;`로 수정해 준다.

5. DB에 dump 파일로 data를 복구한다.

- [Database] MariaDB & MySQL - 데이터베이스 백업 및 복구 방법

Ref

도봉산핵주먹, mariadb, CharacterSet, Encoding, utf8mb4, Migration

From:
http://125.132.25.164/dokuwiki/ - **문제를 잘 정의하는 것은 문제를 절반 해결한 것이다. - 2023.12**

Permanent link:
http://125.132.25.164/dokuwiki/doku.php?id=wiki:mariadb:latin1%EC%97%90%EC%84%9C_utf8mb4%EB%A1%9C_%EB%B3%80%ED%99%98%ED%95%98%EA%B8%B0&rev=1587525128

Last update: **2022/03/10 19:52**