latin1에서_utf8mb4로_변환하기

- description : Table Encoding 변환
- 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. dump 파일을 밀어 넣는다.
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
```

Ref

- 문자 집합(Character Set)과 인코딩(Encoding)

도봉산핵주먹, java, List마지막요소

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=1587519238

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