

utf8mb4와 utf8mb3 차이점은?

MySQL에서 utf8mb4와 utf8mb3의 주요 차이점은 다음과 같습니다:

1. 지원하는 문자 범위

- utf8mb3: 최대 3바이트 길이의 유니코드 문자를 지원합니다. 이모지(emoji)나 일부 특수문자와 같은 4바이트 유니코드 문자는 지원하지 않습니다.
- utf8mb4: 최대 4바이트 길이의 유니코드 문자를 지원합니다. 따라서 이모지, 고대 문자, 특수 기호 등 모든 유니코드 문자를 저장할 수 있습니다^{[1] [2] [3]}.

2. 역사적 배경

- utf8mb3: MySQL에서 초기 utf8 캐릭터셋은 사실상 utf8mb3를 의미했습니다. 이는 유니코드의 기본 다국어 평면(BMP, Basic Multilingual Plane)만 지원하도록 설계되었습니다.
- utf8mb4: 시간이 지나면서 이모지와 같은 4바이트 문자의 필요성이 증가하자 MySQL 5.5부터 utf8mb4가 도입되었습니다. 이는 완전한 UTF-8 지원을 제공합니다^{[2] [3]}.

3. 저장 공간

- utf8mb3: 문자는 1~3바이트로 저장됩니다.
- utf8mb4: 문자는 1~4바이트로 저장됩니다. 따라서 4바이트 문자를 저장할 경우 더 많은 공간을 사용합니다^{[2] [4]}.

4. 호환성 및 권장 사항

- MySQL 8.0에서는 utf8mb3가 더 이상 기본값으로 사용되지 않으며, 향후 제거될 예정입니다. 따라서 새로운 애플리케이션에서는 utf8mb4를 사용하는 것이 권장됩니다^{[2] [3]}.
- utf8mb3는 이전 시스템과의 호환성을 위해 유지되고 있지만, 모든 새로운 개발에서는 utf8mb4를 사용해야 합니다^{[1] [2]}.

5. 설정 예시

데이터베이스 생성 시:

```
CREATE DATABASE my_database
DEFAULT CHARACTER SET utf8mb4
COLLATE utf8mb4_general_ci;
```

테이블 생성 시:

```
CREATE TABLE my_table (
  id INT NOT NULL,
  name VARCHAR(255)
) CHARACTER SET utf8mb4 COLLATE utf8mb4_general_ci;
```

6. 언제 무엇을 선택해야 할까?

- 다국어 데이터나 이모지(emoji)를 다룰 경우: 반드시 utf8mb4를 사용해야 합니다.
- 기존 시스템과의 호환성이 중요한 경우: utf8mb3를 사용할 수 있지만, 가능한 한 업그레이드를 계획해야 합니다.

요약

- utf8mb3: 최대 3바이트 문자만 지원하며, 이모지와 일부 특수문자는 저장할 수 없습니다.
- utf8mb4: UTF-8 표준을 완전히 지원하며, 모든 유니코드 문자를 저장할 수 있습니다.
- 새로운 프로젝트에서는 반드시 utf8mb4를 사용하는 것이 권장됩니다^{[1] [2] [3]}.

✻

1. <https://www.nanumtip.com/qa/267831/>

2. <https://hanainu.tistory.com/48>

3. <https://jay-choe.tistory.com/22>

4. <https://dungbeetle.co.kr/tomcat의-utf-8-과-mysql의-utf8-utf8mb3-utf8mb4-의-차이는-뭘까/>