



8장. URL 단축기 설계

문제 이해 및 설계 범위 확정

긴 url을 짧은 url로 바꿔주는 url 단축기 설계

전체적인 요구사항은 아래와 같다

1. 매일 1억개의 단축 url을 생성할 수 있어야 한다.
2. 단축 url은 짧으면 짧을수록 좋다.
3. 생성된 단축 url은 삭제하거나 갱신할 수 없다.
4. 높은 가용성과 규모 확장성, 장애 감내가 요구된다.

▼ 장애 감내 혹은 내결함성 : 구성하는 하나 이상의 구성요소에 장애가 발생해도 시스템이 중단없이 계속 작동할 수 있는 성질

높은 가용성(HA, High Availability)과 장애 감내(FT, Fault Tolerance) → 비슷하지만
다름

HA는 서비스 중단이 허용되는 최소의 시간이 존재한다. 예를 들어 가용성이 99.99999999999999%인 시스템은 연간 5분의 다운타임을 허용한다. 하지만, FT는 서비스 중단이 전혀 없이 동작해야한다. 비유하자면 스페어 타이어가 있는 자동차는 HA이다. 타이어는 펑크가 날 수 있지만, 빠른 시간내에 타이어를 교체할 수 있다. 반면 엔진이 2개인 쌍발 비행기는 FT이다. 엔진 하나가 고장나도 다른 엔진으로 계속 비행할 수 있다.

HTTP 리디렉션

단축 url 서비스는 단축된 url로 접근하면, 응답 헤더 `Location`에 실제 url을 넣어두고 상태 코드로 301 또는 302를 반환

301 Permanently Moved

- 영구적으로 url이 변경됨을 의미

- 한번 301을 만난 브라우저는 **Location** 값을 캐싱해두고, 단축 url로 접근하면 단축 url 서버에 요청하지 않고 캐싱된 원본 url로 이동
- 단축 url 서버에 부하 줄이고 싶을 때 사용

302 Found

- 일시적으로 url이 변경됐음을 의미
- 별도의 캐싱 없이 항상 단축 url 서버에 실제 요청 보낸 다음, **Location** 헤더를 읽어 원본 url로 이동
- 트래픽 분석 등이 필요할 때 사용

데이터 모델링

원본 url과 단축 url의 쌍을 저장하여 서비스 구현 가능

해시 테이블 써도 되지만, 메모리 용량과 휘발성의 한계 → 관계형 db에 저장해야 한다.

url	
PK	<u>id</u>
	shortURL longURL

shortURL, longURL 컬럼 두 개를 한 row에 저장

URL 단축을 위한 해싱

해시 함수

원래 url을 단축 url로 변환하는 데 사용된다.

해시 값 길이

요구사항 중, 매일 1억개의 단축 url을 생성할 수 있어야 한다.

- 1년에 365억개의 url 생성
- 서비스 운영 기간 10년이면 유일한 url을 3,650억개 이상 생성 필요
- 동시에 '단축 url은 짧으면 짧을수록 좋다' 요구사항 만족해야함

url 구성하는 문자를 알파벳 소문자, 대문자, 그리고 숫자로 제한한다면?

- 사용할 수 있는 문자의 종류 $26 + 26 + 10 = 62$ 개
- $62^6 = \text{약 } 568\text{억}$, $62^7 = \text{약 } 3.5\text{조}$
- 해시 값의 길이가 7 이상이면 요구사항 만족

해시 함수 구현

해시 충돌 후 해소 방법

https://en.wikipedia.org/wiki/Systems_design 을 유명한 해시 함수로 해싱해보자.

- CRC32 : 5cb54054
- MD5 : 5a62509a84df9ee03fe1230b9df8b84e
- SHA-1 : 0eeae7916c06853901d9ccbefbfcaf4de57ed85b

가장 짧은 CRC32도 8글자다.

- 앞 7글자만 사용하는 방법이 있지만, 이렇게 하면 해시 충돌 확률 증가
충돌이 해소될 때 까지 사전에 정한 문자열을 해시값에 덧붙이는 방법 존재
- 해시 충돌 발생할 때마다 db에 쿼리 해야하므로 오버헤드가 큼
 - db 대신 bloom필터를 사용하면 성능 높이기 가능

▼ bloom 필터

어떤 집합에 특정 원소가 있는지 검사할 수 있도록 하는, 확률론에 기초한 공간 효율이 좋은 기술

base-62 변환법

url에 들어갈 수 있는 문자 종류 62개 → 62진법 사용

pk id값을 62진법으로 바꿈

- 만약 PK가 12345678라면, 62진법으로는 `pnfq` 이다.
- 따라서, 단축 URL은 `https://hudiurl.com/pnfq` 가 된다.

두 접근법 비교

해시 후 충돌 해소	base-62 변환
단축 URL의 길이가 고정된다.	가변적인 단축 URL 길이를 갖는다. ID 값이 커지면, URL 길이도 길어진다.
유일성 보장 ID 생성기가 필요없다.	ID를 기반으로 URL이 결정되므로, 유일성이 보장되는 ID 생성기가 필요하다.
충돌을 해소하는 작업이 필요하다. 이로 인한 오버헤드가 발생한다.	충돌이 아예 불가능하다. 유일한 ID로 URL을 생성하기 때문이다.
다음에 생성된 단축 URL을 유추할 수 없다.	ID가 Auto Increment로 생성된다는 가정하에, 다음 URL을 예측할 수 있으므로 보안상 문제가 될 수도 있다.

References

<https://hudi.blog/system-design-interview-alex-xu-8/>