

가상 면접 사례로 배우는
대규모 시스템 설계 기초

이정민

7. 분산 시스템을 위한 유일 ID 생성기 생성

1단계) 문제 이해 및 설계 범위 확정

- ID는 유일해야 함
- ID는 숫자로만 구성되어야 함
- ID는 64비트로 표현될 수 있어야 함
- ID는 발급 날짜에 따라 정렬 가능해야 함
- 초당 10,000개의 ID를 만들 수 있어야 함

2단계) 개략적 설계안 제시 및 동의 구하기

■ 분산 시스템에서 유일성이 보장되는 ID 설계 방안

- 다중 마스터 복제(multi-master replication)
- UUID(Universally Unique Identifier)
- 티켓 서버
- 트위터 스노플레이크(twitter snowflake) 접근법

2단계) 개략적 설계안 제시 및 동의 구하기

■ 다중 마스터 복제

○ 데이터베이스의 `auto_increment` 기능 활용

- 테이블의 특정 컬럼에서 자동으로 숫자를 증가시킴
- 다음 순번 계산 시 k 만큼 증가 (k = 현재 사용 중인 DB수)

○ 단점

- 여러 데이터 센터에 걸쳐 규모를 늘리기 어려움
- ID 유일성은 보장되나, 그 값이 시간 흐름에 따르지 않음
- 서버 추가/제거 시 잘 동작하게 하기 어려움

2단계) 개략적 설계안 제시 및 동의 구하기

■ UUID

○ 유일성이 보장되는 간단한 ID 생성 방법 (128비트)

- 123e4567-e89b-12d3-a456-426614174000
(하이픈으로 구분된 5개의 16진수 문자열)
- low time + mid time + mid time+version + clock sequence & variant + node
- 충돌 가능성이 매우 낮음 (초당 100만개 100년 생성 시, 약 0.00009%)

○ 장점

- 서버 간 조율 없이 독립적 생성가능 (동기화 불필요)
- 각 서버가 ID를 독립적 생성하여 규모 확장 쉬움

○ 단점

- ID가 128비트로 길고, 시간 순으로 정렬 할 수 없음
- ID에 숫자 아닌 값이 포함 될 수 있음

■ 참고) UUID 버전 별 특징

- 1 : 현재시간, **MAC**주소 - 유일성 보장, 보안취약
- 2 : 1버전 유사, **POSIX UID** 사용 - 거의 미사용
- 3 : 이름/네임스페이스 **MD5** 해시 - 보안성 높음
- 4 : 랜덤한 값 기반 - 보안성 높고, 생성 속도 빠름
- 5 : 3버전 유사, **SHA-1** 해시 - 보안 취약

2단계) 개략적 설계안 제시 및 동의 구하기

■ 티켓 서버

- `auto_increment` 기능을 갖춘 데이터베이스 서버(티켓 서버)를 중앙 집중형으로 하나만 사용
- 장점
 - 유일성이 보장되는 숫자로만 구성된 ID를 쉽게 만들 수 있음
 - 구현하기 쉽고, 중소 규모 애플리케이션에 적합함
- 단점
 - 티켓서버가 **SPOF** 되고, 장애 발생 시 모든 시스템이 영향을 받음
 - 이를 해결하기 위해 티켓 서버를 여러대 준비하면 데이터 동기화 문제가 새로 발생함

2단계) 개략적 설계안 제시 및 동의 구하기

■ 트위터 스노플레이크 접근법

○ ID 구조

- 사인 비트 (1) : 나중을 위해 유보
- 타임스탬프 (41) : 기원 시각(epoc) 이후로 몇 밀리초가 경과했는 지 나타내는 값
- 데이터센터 ID (5) : 32개의 데이터센터를 지원 할 수 있음
- 서버 ID (5) : 데이터 서버당 32개 서버를 사용할 수 있음
- 일련번호 (12) : 각 서버에서 ID를 생성할 때마다 이 일련번호를 1만큼 증가 시킴, 1밀리초가 경과할 때마다 0으로 초기화

3단계) 상세 설계

■ 타임스탬프

- ID가 시간 순으로 정렬 가능한 이유
- 타임스탬프의 최대 값 : 약 **69년 (41비트)**
- 타임스탬프 계산
 - 타임스탬프 = UTC 밀리초 - 트위터 기원시각(epoch)

■ 일련번호

- 일련번호의 최대 값 : **4096개 (12비트)**
- **1 밀리초** 이내 동시에 생성 시 **0**이상 값으로 증가

4단계) 마무리

■ 추가 논의 사항

- 시계 동기화(clock synchronization) : NTP(Network Time Protocol) 등
- 각 섹션 길이 최소화
 - 예) 동시성이 낮고 수명이 긴 애플리케이션의 경우 일련번호 섹션 길이를 줄이고 타임스탬프 섹션의 길이를 늘리는 듯
- 고가용성 : ID 생성기는 **Mission Critical** 컴포넌트이므로 고가용성을 제공해야 함