

4주차

5장: 안정 해시 설계

안정 해시

- 안정 해시는 해시 테이블 크기가 조정될 때 평균적으로 오직 k/n 개의 키만 재배치하는 해시 기술
 - k: 키의 개수
 - n: 슬롯(slot)의 개수
- 대부분 전통적 해시 테이블은 슬롯의 수가 바뀌면 거의 대부분의 키를 재배치

해시 공간과 해시 링

- 해시 함수 f로는 SHA-1를 사용한다고 가정
- 이 함수의 출력 값 범위는 x_0, \dots, x_n 과 같음
- SHA-1의 해시 공간(hash space) 범위는 0부터 $2^{160} - 1$ 까지라고 알려져 있음
- 해시 공간의 양쪽을 구부려 접으면 해시 링이 만들어짐

해시 서버

- 해시 함수 f를 사용하면 서버 IP나 이름을 링 위의 어떤 위치에 대응시킬 수 있다.

해시 키

- 나머지 연산 % 은 사용하지 않는 해시 함수 사용
- 키도 해시 링 위의 어느 지점에 배치

서버 조회

- 어떤 키가 저장되는 서버는, 해당 키의 위치로부터 시계 방향으로 링을 탐색해 나가면서 만나는 첫 번째 서버

서버 추가

- 서버를 추가하더라도 키 가운데 일부만 재배포하면 됨
- 예시로 서버 4가 추가되면 key0만 재배포됨
- 서버 4가 추가되면 key0은 서버4에 저장되게 됨

서버 제거

- 하나의 서버가 제거되면 키 가운데 일부만 재배포 됨

기본 구현법의 두가지 문제

- 서버가 추가되거나 삭제되는 상황을 감안하면 **파티션의 크기를 균등하게 유지하는 게 불가능**
 - 파티션이란 인접한 서버 사이의 해시 공간
 - s2의 파티션이 거의 두 배로 커지는 상황
- **키의 균등 분포**를 달성하기가 어려움
 - 이를 해결하기 위해 가상 노드 또는 복제라 불리는 기법을 사용

가상 노드

- 실제 노드 또는 서버를 가리키는 노드로서, 하나의 서버는 링 위에 여러 개의 가상 노드를 가질 수 있음
- 각 서버는 하나가 아닌 여러 개 파티션을 관리해야 한다.
- 키 위치로부터 시계 방향으로 링을 탐색하다 만나는 최초의 가상 노드가 해당 키가 저장될 서버
- 가상 노드의 개수를 늘릴수록 키의 분포는 점점 더 균등해짐
- 표준 편차가 작아져서 데이터가 고르게 분포되기 때문
- 가상 노드의 개수를 늘리면 표준 편차가 줄어들지만 저장할 공간이 늘어나니 트레이드 오프가 필요

안정 해시의 이점

- 서버가 추가되거나 삭제될 때 재배포되는 키의 수가 최소화
- 데이터가 보다 균등하게 분포하게 되므로 수평적 규모 확장성을 달성하기 쉬움
- 핫스팟(hotspot) 키 문제를 줄임.

- 특정 샤드에 대한 접근이 지나치게 빈번하면 서버 과부하 문제가 생길 수 있는데, 안정해시는 데이터를 좀 더 균등하게 분배하므로 이런 문제 줄인다.