

01.09

해시 충돌

- hash는 단방향의 암호화 알고리즘이다. hashmap과 hashtable은 hash알고리즘을 이용한 자료구조이다!

- open addressing

- 1) 다음칸 탐색

- open addressing에서 해시 충돌 시 n칸을 건너뛴 다음 버킷에 저장하는 경우 데이터들이 특정 위치에만 밀집할 수 있다.(clustering)

- 2) 제곱 탐색

- 데이터들이 특정 위치에 밀집하는 문제를 어느정도 해결하기 위해 나옴

- 3) 이중 해시

- 그럼에도 불구하고 Hash값이 동일하다면 건너뛴 칸의 개수도 똑같이 때문에, 똑같이 N번의 탐색을 할 수 있고 clustering이 여전히 발생해서 나옴.

- sepearte chaining

- 1) linkedList

- Hash Function으로 index의 위치를 찾는데 $O(1)$ 의 시간 복잡도가 걸린다고 해도, index의 위치에서 값을 가져올 때 해당 값을 찾기 위해 또 리스트를 하나씩 탐색하는 과정을 거쳐야 해서 $O(N)$ 의 시간 복잡도가 나온다.

- 2) tree

- linkedList대신 트리자료를 사용함으로써 시간복잡도 단축시켰다.

안정해시설계 - 수평적 규모 확장성 달성하기

- 안정적으로 데이터 균등 분배를 위해 사용되었다. 해시테이블 크기가 조정될때 평균적으로 K/n 개의 키만 재배치하는 것.

- 서버와 키를 균등분호 해시함수를 이용해 해시링에 배치
- 키의 위치에서 링을 시계방향으로 탐색하다 만나는 최초의 서버가 키가 저장될 서버

장점

- 서버가 추가되거나 삭제될때 재배치되는 키의 수가 최소화됨
- 데이터가 보다 균등 분포(수평적 규모 확장성 달성)

단점

- 파티션 크기 균등 유지 X
- 키의 균등 분포 X

극복 : 가상 노드 기법 도입 → 핫스팟 키문제 감소 - 서버 과부하 감소

하나의 서버가 여러개의 가상노드를 갖는것.