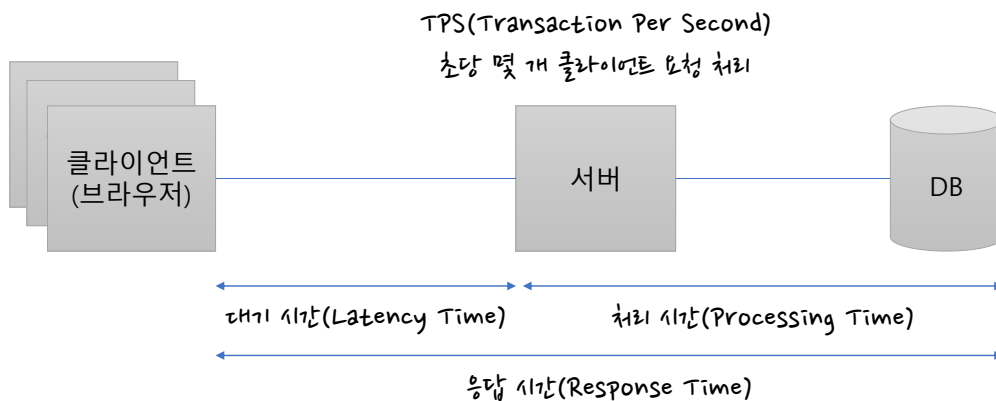


# 초식 : 초보자용 웹 서비스 성능 올리기

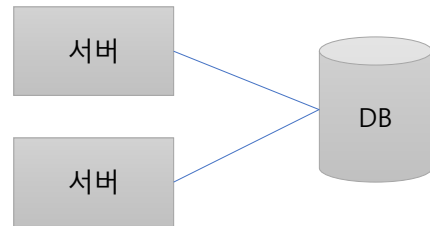
최범균

## 서버 성능 기본 지표 2개 : 응답 시간, 처리량



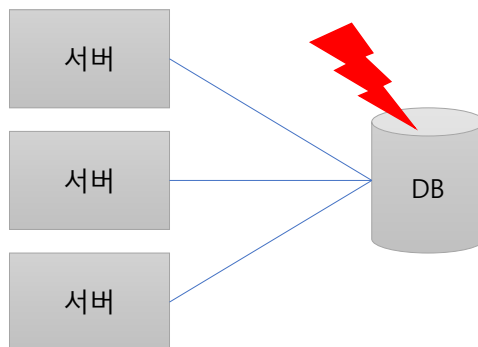
## TPS를 높이려면

- 가장 쉬운 방법 : 서버 늘리기
  - 단순하게 서버 1대가 10 TPS면, 2대면 20 TPS
- 비슷한 방법으로 스레드 풀+DB 커넥션 풀 늘리기
  - 동시 처리할 수 있는 개수
    - 스레드 풀 5, 처리 시간 1초 → TPS 5
    - 스레드 풀 10, 처리 시간 1초 → TPS 10



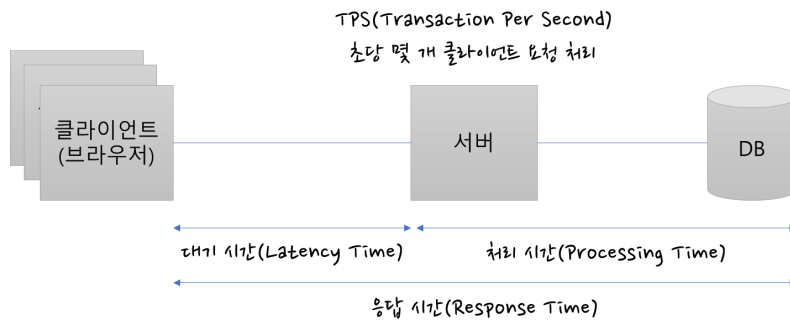
## 서버 늘리기/스레드 풀 늘리기 한계

- DB에 대한 부하가 임계치를 넘어가면 먹통 증상 발생
  - DB 쿼리 시간 증가 → 처리 시간 길어짐 → TPS 떨어짐



## TPS를 높이려면

- 기본적으로 처리 시간을 줄여야 함
  - 처리시간 1초, 스레드풀 10 → 10 TPS
  - 처리시간 0.5초, 스레드풀 10 → 20 TPS

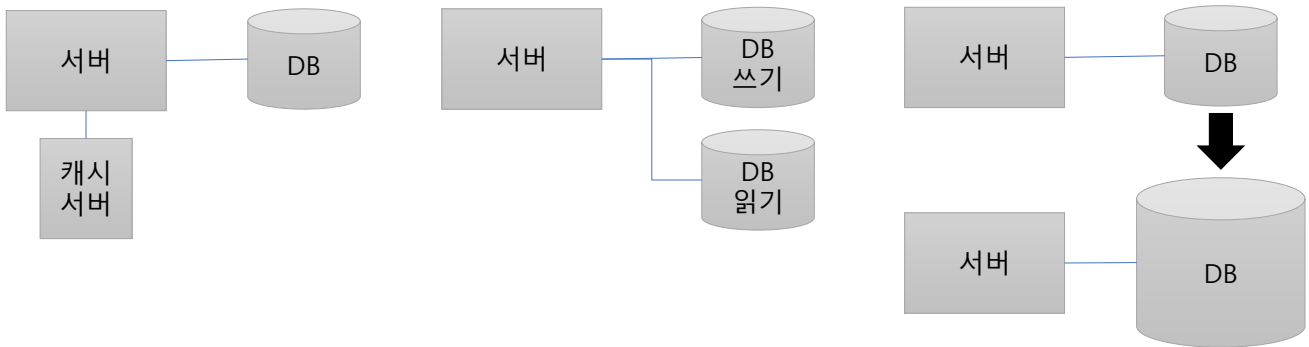


## 처리 시간을 줄이려면

- 처리 시간에서 비중이 높은 대상을 찾아 줄여야 함
- 비중이 높은 대상
  - DB 연동
  - API 호출
  - 데이터 집계/계산

## 처리 시간 줄이기 : DB

- 보통 세 가지 방법 사용: 쿼리 튜닝, 캐시, 장비빨
- 장비빨: 1 Primary/Replica읽기 전용 DB, 2 하드웨어 업그레이드



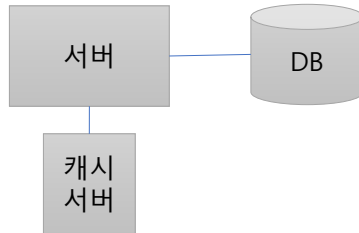
## 처리 시간 줄이기 : API 호출

- API 호출 응답 시간 증가 → 처리 시간 증가
- API 호출 성능 개선 방법 2가지: 캐시, 호출 제거
  - 호출 제거 : 메시징 + 비동기 연동



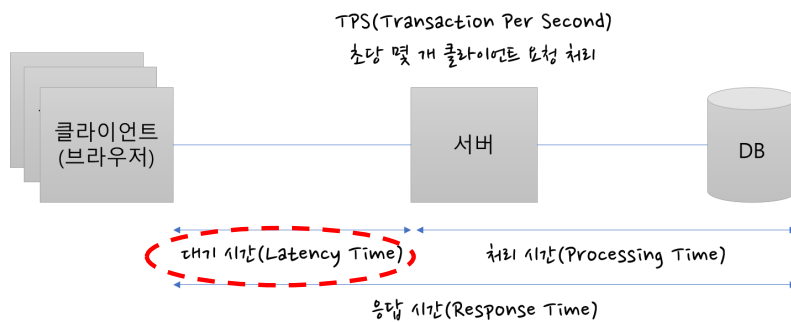
## 처리 시간 줄이기 : 데이터 집계/계산

- 미리 계산해서 캐시나 DB에 보관하기
  - 통계류(예, 좋아요 수)



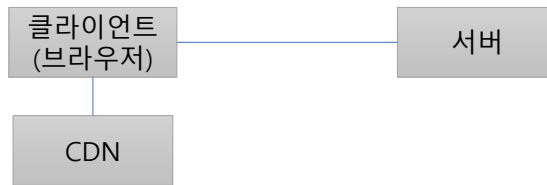
## 응답 시간을 줄이려면

- 대기 시간도 줄일 필요 있음



## 대기 시간 줄이기 : 대역폭

- 대역폭이 작으면 클라이언트 개수가 증가할 때 주고 받는 속도가 급격히 느려짐
  - 고속도로에 차가 증가하면 속도가 느려지는 것과 유사
- 세 가지 방식 고려
  - 응답 크기 줄이기 : 응답 압축, 이미지 파일 크기 줄이기
  - 트래픽 분리하기 : 이미지, 정적 파일을 CDN을 통해 제공
  - 대역폭 늘리기 (비용 측면에서 CDN이 유리)



## 정리

