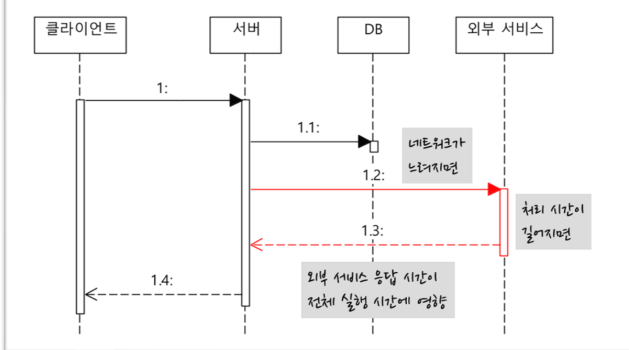


웹 서비스 Resilience

외부 API 장애에 영향 덜 받는 3가지 방법

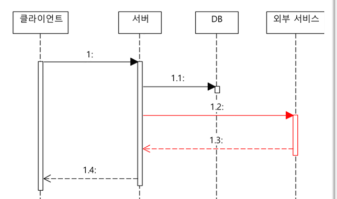
최범균

외부 연동과 성능



외부 서비스의 응답 시간과 성능 저하

- 예: 서버 쓰레드풀 5개, 평균 응답 시간 0.2초
 - 응답 시간에서 외부 연동이 차지하는 시간이 0.1초
 - 1개 쓰레드가 초당 5개 요청 처리
 - 5개 쓰레드가 초당 25개 요청 처리
 - 25 TPS
- 외부 연동 0.1초 → 0.9초 증가하면
 - 응답 시간 1초로 증가
 - 1개 쓰레드가 초당 1개 요청 처리
 - 5개 쓰레드가 초당 5개 요청 처리
 - 5 TPS



1초 2초 3초 4초 5초



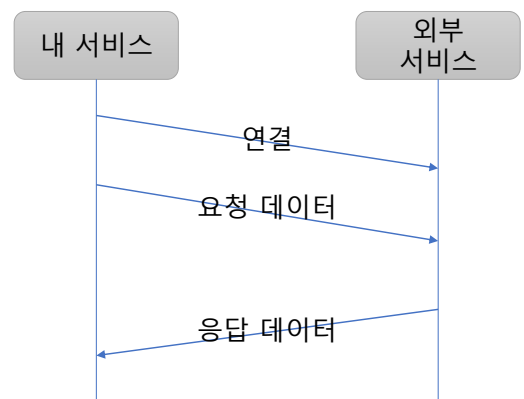
초만에 응답을 받았던 클라이언트가 5초만에 응답을 받음

동기 방식으로 외부 서비스 호출할 때 외부 서비스 장애 영향 덜 받는 3가지 방법

- 타임아웃
- 벌크헤드 Bulkhead
- 서킷 브레이커 Circuit Breaker

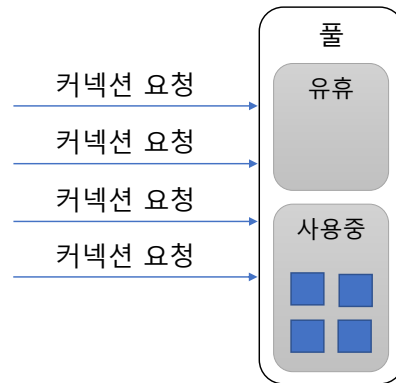
타임 아웃

- 오래 대기하는 것보다
빨리 에러를 발생시키는 것이 나음
 - 동시에 대기하는 외부 연동 개수가
 많아질수록 내 서비스는 점점 먹통/느려짐
- 외부 API 호출시 두 개 타임아웃 설정
 - 연결 시간 타임 아웃
 - 보통 1초~5초 이내로 설정
 - 응답 시간 타임 아웃
 - 보통 1초~3초 사이 적당
 - 외부 서비스 특징에 따라 길게 주기도 함
 - 예: 결제 승인은 30초



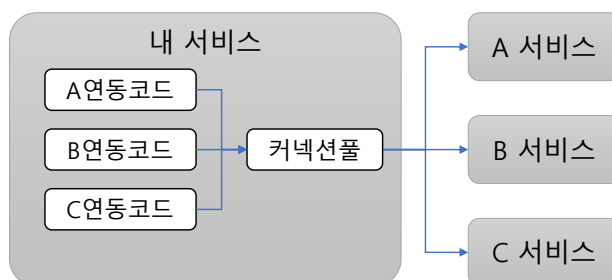
타임 아웃

- HTTP 커넥션 풀 사용시 추가 타임 아웃 설정
 - 풀에서 커넥션 구할 때 대기하는 최대 시간
 - 보통 1초~5초 이내로 설정



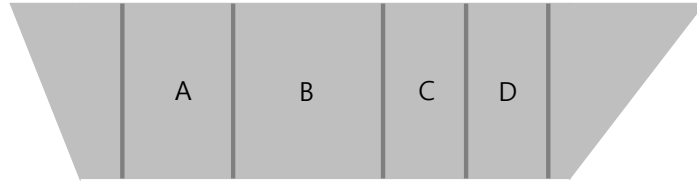
외부 연동 서비스가 많을 때

- 특정 서비스 장애 → 전체 영향
 - 예: A, B, C 연동에 대해 커넥션풀을 공유할 때
 - A 서비스의 장애로 응답 시간 지연 발생하면
 - → 풀에 남은 유휴 커넥션이 줄면서
 - → 풀에서 커넥션을 구하는 대기 시간이 증가하고
 - → B, C 서비스에 대한 연동도 같이 대기하게 됨



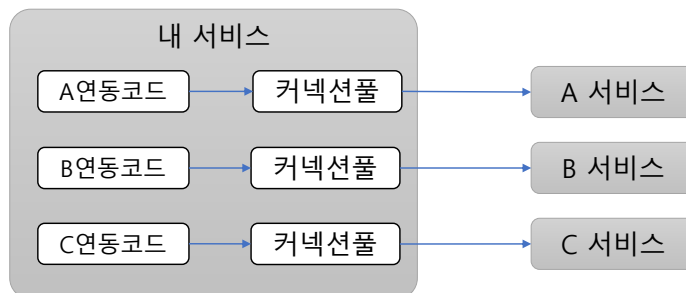
벌크헤드

- 기능/서비스/클라이언트 마다 자원 사용 분리
 - 한 기능/서비스 장애가 다른 기능/서비스에 주는 영향 최소화



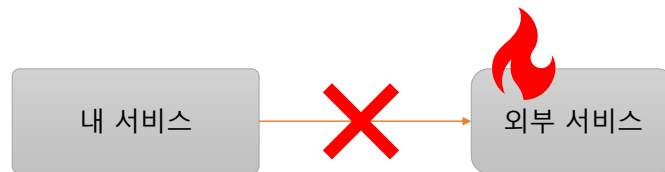
벌크헤드

- 예: 외부 서비스마다 별도 커넥션풀 사용
 - 트래픽 규모에 따라 서로 다른 커넥션 풀 설정
 - A 서비스와 연동이 느려져도 B,C 서비스 연동은 영향 감소



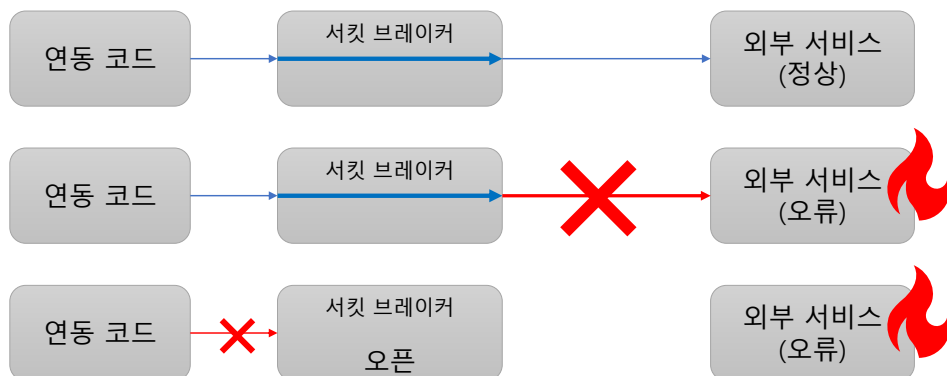
외부 시스템 장애 지속시

- 계속되는 타임아웃/50x 응답에 따른 서비스 에러
- 외부 서비스가 비정상임에도 계속해서 요청 보냄
- 응답 시간 느려짐, 처리량 감소

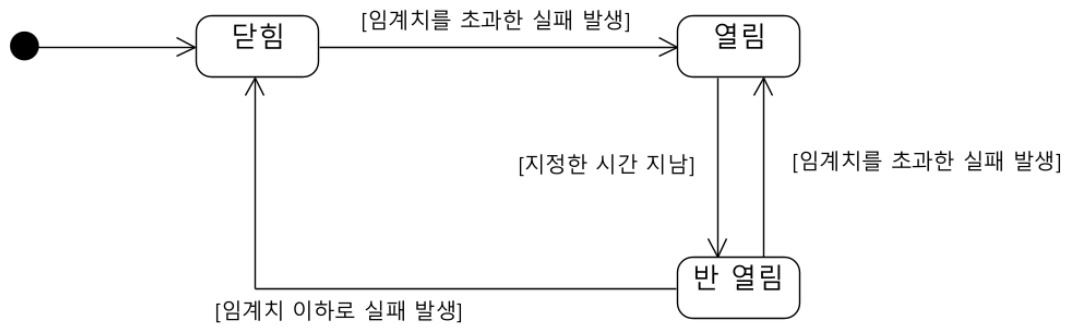


서킷 브레이커

- 오류 지속시 일정 시간동안 기능 실행을 차단함
 - 기능을 실행하지 않고 바로 에러 응답
 - 빠른 실패(fail fast) → 외부 서비스 장애에 따른 응답 시간 증가 감소



서킷 브레이커 동작 방식



정리

- 연동 시스템 장애시 영향을 줄이는 세 가지 방법
 - 타임 아웃
 - 벌크헤드
 - 서킷브레이커