

2장 : 개략적인 규모 추정

④ 난이도	★★
▣ 학습날짜	@2025년 12월 4일

응답지연 측정 값

가용성

요구량 측정 방법 (QPS)

가정

추정

시스템 용량이나 성능 요구사항을 개략적으로 추정해 보아야 할 때 측정하는 방법에 대한 챕터이다.

응답지연 측정 값

- L1 캐시 참조
- 분기 예측 오류
- L2 캐시 참조
- 뮤텍스 락 / 언락
- 주 메모리 참조
- Zippy로 1KB 압축
- 1 Gbps 네트워크로 2KB 전송
- 메모리에서 1MB 순차적으로 read
- 같은 메모리 센터 내에서의 메시지 왕복 지연시간
- 디스크 탐색 (seek)
- 네트워크에서 1MB 순차적으로 read
- 디스크에서 1MB 순차적으로 read
- 한 패킷의 CA로부터 네덜란드까지의 왕복 지연시간

가용성

고가용성은 시스템이 오랜 시간 동안 지속적으로 중단 없이 운영될 수 있는 능력이다.

대부분 고가용성은 숫자 9를 사용해 표시한다.

SLA (Service Level Agreement)는 서비스 사업자가 보편적으로 사용하는 용어로, 서비스 사업자와 고객 사이에 맺어진 합의를 의미한다.

가용률	하루당 장애시간	연간 장애시간
99%	14.40분	3.65일
99.9%	1.44분	8.77시간
99.99%	8.64초	52.60분
99.999%	864ms	5.26분
99.9999%	86.40ms	31.56초

요구량 측정 방법 (QPS)

가정

- 월간 능동 사용자는 3억 명이다.
- 50%의 사용자가 트위터를 매일 사용한다.

- 평균적으로 각 사용자는 매일 2건의 트윗을 올린다
- 미디어를 포함하는 트윗은 10% 정도다.
- 데이터는 5년간 보관된다.

추정

QPS (Query Per Second) 추정치

- 일간 능동 사용자 (Daily Active User, DAU) = 3억 * 50% = 1.5억
- QPS = 1.5억 * 2 트윗 / 24시간 / 3600초 = 약 3500
- 최대 QPS (Peak QPS) = 2 * QPS = 약 7000

미디어 저장을 위한 저장소 요구량

- **평균 트윗 크기**
 - tweed_id : 64B
 - 텍스트 : 140B
 - 미디어 : 1MB
- **미디어 저장소 요구량** : 1.5억 * 2 * 10%
- 5년간 미디어를 보관하기 위한 저장소 요구량 : 30TB * 365 * 5 = 약 55PB