

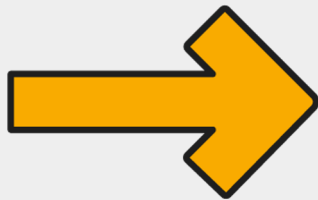


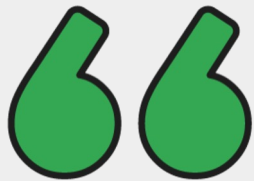
Google Developer Group
Incheon National University

가상 면접 사례로 배우는 대규모 시스템 설계 기초

개략적인 규모 추정

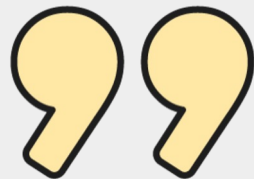
Back-of-the-envelop estimation





보편적으로 통용되는 성능 수치상에서
사고 실험(thought experiments)을
행하여 추정치를 계산하는 행위

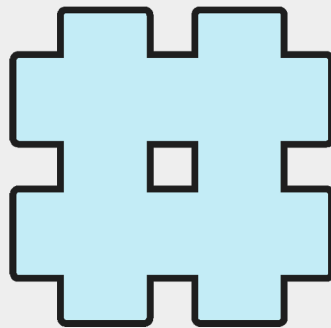
- Senior Fellow Jeff Dean



2의 제곱수

데이터 볼륨 단위

2의 x 제곱	근사치	이름	축약형
10	1천(thousand)	1킬로바이트(Kilobyte)	1KB
20	1백만(million)	1메가바이트(Megabyte)	1MB
30	10억(billion)	1기가바이트(Gigabyte)	1GB
40	1조(trillion)	1테라바이트(Terabyte)	1TB
50	1000조(quadrillion)	1페타바이트(Petabyte)	1PB



응답지연 값



시스템에서 요청이 처리되는 데 걸리는 시간

=> 성능 추정 시, 각 단계의 응답 지연을 고려하는 것이 중요

Ex. 디스크 탐색(seek) - 10ms

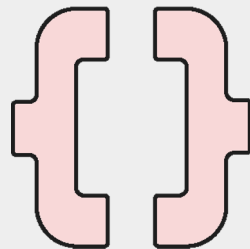
하드 디스크(HDD)에서 원하는 데이터를 읽거나 쓰기 위해
데이터를 저장한 위치로 디스크 헤드가 움직이는 과정



100,000

RAM이 100,000배 빠름 (캐시 활용 필수)

가용성에 관계된 수치들



고가용성(high availability)

: 시스템이 오랜 시간 동안 지속적으로 중단 없이 운영될 수 있는 능력

가용률	하루당 장애시간	주당 장애시간	개월당 장애시간	연간 장애시간
99%	14.40분	1.68시간	7.31시간	3.65일
99.9%	1.44분	10.08분	43.83분	8.77시간
99.99%	8.64초	1.01분	4.38분	52.60분
99.999%	864.00밀리초	6.05초	26.30초	5.26분
99.9999%	86.40밀리초	604.80밀리초	2.63초	31.56초

SLA(Service Level Agreement)

: 서비스 제공자(service provider)가 보장하는 최소 성능 및 가용성 기준

QPS와 저장소 요구량 추정

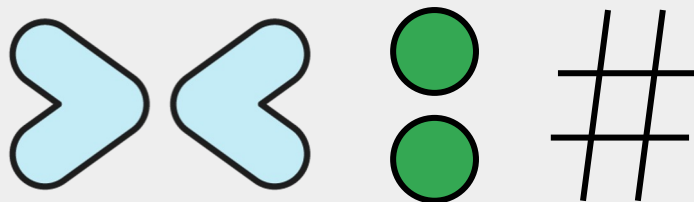
- 초당 요청 수 (Queries Per Second, QPS)

시스템이 처리해야 할 부하

- 저장소 요구량 (Storage Estimation)

데이터 저장 용량을 추정하는 방법

(데이터 크기 * 기간)



개략적인 규모 추정을 할 때:

1. 2의 제곱수를 이용해 용량과 성능을 빠르게 추정
2. 응답 지연 값을 참고하여 성능 병목을 찾기
3. SLA와 가용성을 고려하여 다운타임을 최소화
4. QPS를 계산하여 서버 부하를 예측
5. 저장소 요구량을 계산하여 인프라를 설계

