

### 3

## 230920 2. 계략적인 규모 추정

### ▼ 2. 계략적인 규모 추정



계략적인 규모 추정은 보편적으로 통용되는 성능 수치상에서 사고 실험을 행하여 추정치를 계산하는 행위로서, 어떤 설계가 요구사항에 부합할 것인지 보기 위한 것

- 구글의 시니어 펠로 제프 딘

시스템 설계 면접을 볼 때, 시스템 용량이나 성능 요구사항을 새략적으로 추정해 보라는 요구를 받게 됨

계략적 규모 추정을 효과적으로 해 내려면 규모 확장성을 표현하는 데 필요한 기본기에 능숙해야 함

### ▼ 2의 제곱수

분산 시스템에서 다루는 데이터 양은 엄청나게 커질 수 있음

하지만 그 계산법은 기본을 크게 벗어나지 않음

따라서 제대로 된 계산 결과를 얻으려면 데이터 볼륨의 단위를 2의 제곱수로 표현하면 어떻게 되는지 알아야 함

최소 단위 ⇒ 1바이트 : 8비트

ASCII 문자 하나 ⇒ 1바이트

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

구글의 제프 딘이 공개한 컴퓨터에서 구현된 연산들의 응답지연값

빠른 컴퓨터가 등장하면서 더이상 유효하지 않는 값들도 있지만, 컴퓨터 연산들의 처리 속도가 어느 정도인지 짐작할 수 있게 해줌

수치들의 분석 결과

- 메모리는 빠르지만 디스크는 아직도 느리다
- 디스크 탐색은 가능한 한 피하라
- 단순한 압축 알고리즘은 빠르다
- 데이터를 인터넷으로 전송하기 전에 가능하면 압축하라
- 데이터 센터는 보통 여러 지역에 분산되어 있고, 센터들 간에 데이터를 주고받는 데는 시간이 걸린다

## ▼ 가용성에 관계된 수치들

### 고가용성

시스템이 오랜 시간 동안 지속적으로 중단 없이 운영될 수 있는 능력

퍼센트로 표현(100% → 시스템이 단 한 번도 중단된 적 없음)

대부분의 서비스는 99% ~ 100%의 고가용성을 유지함

### SLA

서비스 사업자가 보편적으로 사용하는 용어

서비스 사업자와 고객 사이에 맺어진 합의

### 가용시간

서비스 사업자가 제공하는 서비스의 가용시간이 공식적으로 기록되어 있음

아마존, 구글, 마이크로소프트 같은 사업자는 99% 이상의 SLA를 제공

관습적으로 숫자 9를 사용해 표시

9가 많을수록 좋음

### 9의 개수와 시스템 장애 시간의 관계

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

## ▼ 면접 팁

문제를 풀어나가는 절차가 가장 중요

올바른 절차를 밟느냐가 결과를 내는 것보다 중요

면접관은 문제 해결 능력을 보고 싶어 함

### 1. 근사치를 활용한 계산

- 면접장에서 복잡한 계산을 하는 것은 어려움
- 계산 결과의 정확함을 평가하는 것이 아님
- 적절한 근사치를 활용하여 시간 절약

### 2. 가정들은 적어 두기

- 나중에 살펴볼 수 있도록

### 3. 단위를 붙여라

- 단위를 붙이는 습관을 들여 모호함 방지

### 4. 많이 출제되는 개략적 규모 추정 문제

- QPS
- 최대 QPS
- 저장소 요구량
- 캐시 요구량
- 서버 수

## ▼ 토론

시스템의 가용성을 높이기 위해서는 추가적인 비용이 많이 드는데 이때 이익과 비용에 대한 균형을 어떻게 유지하는 것이 좋을지

- 가용성에 대한 레벨을 결정해야 함
- 어떤 비즈니스 프로세스는 가용성을 더 중요시하게 생각할 수도 있고, 다른 경우는 적절한 수준의 가용성으로 충분할 수도 있음
- 고가용성
  - 금융 거래

- 의료 기기
- 항공 교통
- 영상 스트리밍
- 적절한 수준
  - 개인 블로그
  - 게임 애플리케이션
    - 일부 게임 애플리케이션 서버는 상호작용이 매우 빈번하고 대규모 다중 플레이어 온라인 게임을 호스팅합니다. 가용성은 중요하지만, 모든 게임에 대한고가용성을 구현하는 것은 비용이 많이 들기 때문에 일부 덜 중요한 게임에서는 적절한 수준의 가용성을 제공