Back of the Envelope Estimation

- Power of 2
  - A byte is a sequence of 8 bits.
  - ASCII character uses one byte of memory.

Power	Approximate value	Full name	Short name
10	1 Thousand	1 Kilobyte	1 KB
20	1 Million	1 Megabyte	1 MB
30	1 Billion	1 Gigabyte	1 GB
40	1 Trillion	1 Terabyte	1 TB
50	1 Quadrillion	1 Petabyte	1 PB

- Latency numbers every programmer should know

Operation name	Time
L1 cache reference	0.5 ns
Branch mispredict	5 ns
L2 cache reference	7 ns
Mutex lock/unlock	100 ns
Main memory reference	100 ns
Compress 1K bytes with Zippy	10,000 ns = 10 µs
Send 2K bytes over 1 Gbps network	20,000 ns = 20 µs
Read 1 MB sequentially from memory	250,000 ns = 250 µs
Round trip within the same datacenter	500,000 ns = 500 µs
Disk seek	10,000,000 ns = 10 ms
Read 1 MB sequentially from the network	10,000,000 ns = 10 ms
Read 1 MB sequentially from disk	30,000,000 ns = 30 ms
Send packet CA (California) ->Netherlands->CA	150,000,000 ns = 150 ms

- Conclusions
  - Memory is fast but the disk is slow.
  - Avoid disk seek if possible.
  - Simple compression algorithms are fast.
  - Compress data before sending it over the internet if possible.
  - Data centers are usually in different regions, and it takes time to send data between them.

- Availability Numbers

  - High availability is the ability to be continuously operational for a long time.
  - Availability is measured in percentage.
  - Most services fall between 99% and 100%
  - Uptime is measured in nines.

Availability %	Downtime per day	Downtime per year
99%	14.40 minutes	3.65 days
99.9%	1.44 minutes	8.77 hours
99.99%	8.64 seconds	52.60 minutes
99.999%	864.00 milliseconds	5.26 minutes
99.9999%	86.40 milliseconds	31.56 seconds

- Example : Estimate Twitter QPS and storage requirement

  - Assumption
    - 300 million monthly active users.
    - 50% of users use Twitter daily.
    - Users post 2 tweets per day on average.
    - 10% of tweets contain media.
    - Data is stored for 5 years.

  - Estimations

    - Query per second (QPS) estimation :
      - Daily Active Users (DAU) = 50% of 300 million
        = 150 million
      - Tweets QPS = 150 million × 2 tweets / 24 hour / 3600 sec
        = ~3500
      - Peak QPS = 2 × QPS = ~7000

  - Storage Estimates
    - Average tweet size
      - tweet_id - 64 bytes
      - text - 140 bytes
      - media - 1 MB
    - Media Storage : 150 million × 2 × 10% × 1 MB
      = 30 TB per day
    - 5 year media storage : 30 TB × 365 × 5
      = ~55 PB

- Tips

  - Rounding and Approximation
  - Write down all assumption
  - Label your units
  - Commonly asked estimations - QPS, Peak QPS, storage, cache, number of servers etc.