System Design Interview Cheatsheet

prepared by Mark Klenk (ex-Google EM/IGotAnOffer Coach)

Data Considerations

- ❑ Types (e.g. strings, ids, blobs)
- ❑ Sizes (e.g. 1KB average, GBs)
- ❑ Persistence (ephemeral vs. permanent)
- ❑ Access Patterns (e.g. reads & writes)
- ❑ *Sensitivity (e.g. SPII, PII, public)*
- ❑ *Placement (e.g. GDPR)*

Scale Considerations

- ❑ Storage (content & metadata)
  - Bytes
  - Number of DB entries / rows
- ❑ Traffic (writes / ingress & reads / egress)
  - Queries per second/QPS (proxy for CPU)
  - Bandwidth

Scale Units

Magnitude	Suffix	Prefix	Storage
10^3 (1,000)	thousand	kilo	KB(ytes)
10^6	million	mega	MB
10^9	billion	giga	GB
10^12	trillion	tera	TB
10^15	quadrillion	peta	PB
10^18	quintillion	exa	EB
10^21	sextillion	zetta	ZB

Object Sizes

- Key / Attribute: 100B - 10KB
- Document / Photo / MP3: 1-10MB
- Phone / Web Video: 10-100MB
- HD Video / Movie: 1GB - 10GB

Storage Estimation

❏ **Calculate for 3 years (approx. 1,000 days)**

❏ **Number of objects =**
<daily writes> * 1,000 [days]
TIP: choose the next highest scale unit
Example: 10 billion writes / day → 10 trillion objects over 3 years

❏ **Total Bytes =**
<daily writes> * <object size> * 1,000 [days]
TIP: use exponent math and next highest scale unit
Example: 10 billion writes / day, 1KB objects: 10^10 * 10^3 = 10^(10+3) = 10^13 = 10TB / day → 10PB over 3 years

Traffic Estimation

❏ **Assume 100,000 seconds / day, apply peak factor**
❏ **Objects < 10KB → Qps | 1MB+ → Bandwidth**
❏ **QPS (aka TPS / RPS) = <daily API calls> / 100,000**
Example: 100 billion reads / day = 10^10 / 10^5 = 10^5 → 100,000 reads / sec average → 500,000 reads / sec peak

❏ **Bandwidth [bits / second] = <daily API calls> * <object size> * 10 [bits / byte] / 100,000**
TIP: use exponent math
*Example: 1 billion writes / day, 10MB objects: 10^9 * 10^7 * 10^1 / 10^5 = 10^(9+7+1-5) = 10^12 = 1Tbit/sec ave → 3Tbit/sec peak*