**1st round**：

user🡺 API gateway (Routing, Authentication, Authorization, rate limiting)

API 写明功能

SQL DB **ACID** property; Non-SQL, MongoDB; **Object Storage** (AWS S3)

**2nd round**: Check Latency

加 Database Index; 加 Cache (Redis key-value DB), eviction policies: LRU or LFU

要不要加 CDN?

**3rd round**: how to scale:

讨论需不需要 divide API?(need api gateway);

Horizontal Scaling API (load balancer—consistency hashing or Rendezvous hashing)

Config: Scaling setup: CPU usage stays above 70%, add 1 server; below 30%, remove 1 server. set a **cooldown period**, 5 mins

**4th round**: check Availability

Check 999 for SLA; Snapshot for DB?  
Redundancy for DB (active or passive) (azure Geo-Replication)

要不要加：  
**message queue**: 处理拥挤，提供 high availability

**Rate limiting** 一般是加在 API gateway

**log server** (Time series DB) + Monitor to setup actions to do when trigger alert

**TTL**: Time To Live **cron job**: time + actions