HLD Decision Flowchart (Scale → Component Choice)

Estimate Scale	
Users, QPS, R/W Ratio, Storage, Bandwidth	↓
Database Choice	
<10K QPS, <1TB → SQL (MySQL/Postgres) Read-heavy → SQL + Replicas + Cache Write-heavy → NoSQL (Cassandra, DynamoDB) Flexible schema → MongoDB Graph → Neo4j	↓
Caching Layer	
Reads >> Writes → Redis/Memcached Global/Static → CDN Eviction: LRU, LFU, TTL	↓
Storage	
Small files → DB Large media → Blob/Object (S3, GCS) Global delivery → CDN	↓
Traffic Management	
<10K QPS → Single LB 10K–100K QPS → LB + Cache >100K QPS → GSLB + Sharding API Gateway → Auth, Rate Limit	↓
Async & Messaging	
Background jobs → RabbitMQ High-scale stream → Kafka/Pulsar Dead-letter → Retry handling	↓
Extra Components	
Search → Elasticsearch Monitoring → Prometheus + Grafana Logging → ELK Stack Security → OAuth/JWT, Rate-limit	End