

System Scalability Analysis Worksheet

Fill this during mock interviews to estimate load, spot bottlenecks, and justify scaling choices.

Name: _____

Date: _____

1) Traffic Patterns

Inputs

- Read/Write ratio (e.g., 100:1): _____
- Avg RPS: _____ Peak multiplier (x3-x5): _____
- Peak RPS: _____ Burst duration: _____
- Latency target (p95/p99): _____
- Availability target (e.g., 99.9%): _____

2) Scaling Dimensions

Capacity checklist

- Compute (CPU/threads) Notes: _____
- Storage (size/IOPS) Notes: _____
- Network (egress) Notes: _____
- Database (QPS/locks) Notes: _____
- Cache (hit rate/RAM) Notes: _____

3) Bottleneck Identification

Where does it break first at 10x? (be specific)

Component: _____ Why: _____
Symptom (timeouts, CPU, saturation, queue lag): _____
Evidence/metric to watch: _____

4) Scaling Strategies

Pick the smallest change that removes the bottleneck Horizontal vs Vertical

- Horizontal (add instances) ■ Vertical
(bigger box)
Notes: _____

Common Levers

- Caching layer (what + TTL + invalidation)
- Data partitioning/sharding (key + routing)
- Async queues (what + idempotency)
- Read replicas / CQRS / denorm

5) Trade-offs Analysis

Explain the cost of your choice (keep it crisp)

Latency: _____ Cost: _____ Complexity: _____ Consistency: _____

Risk/Failure modes introduced: _____

Mitigation/monitoring: _____

When would you revisit this decision? _____