

Rate Limiter Interview Prep Checklist

Complete Walkthrough Guide

Phase 1: Requirement Clarification (3-5 minutes)

- What's the rate limit identifier? (user ID, IP, API key)
- What scale? (RPS: thousands vs millions)
- Are bursts acceptable?
- Limit exceeded behavior? (reject, queue, throttle)
- Required accuracy? (exact vs approximate)
- Read vs write vs both?

Phase 2: High-Level Design (5-7 minutes)

- Sketch request flow from client to backend
- Identify where limiter sits (gateway vs service)
- Show three components: interceptor, decision engine, data store
- Explain what data flows through system

Phase 3: Deep Dive (10-15 minutes)

- Choose algorithm based on requirements (fixed window, sliding window, token bucket, leaky bucket)
- Design Redis schema (key pattern, data stored, TTL)
- Define API contract (request headers, response codes, rate limit headers)
- Discuss scaling approach (local cache + shared store)

Phase 4: Trade-Offs & Variations (5-10 minutes)

- Accuracy vs performance trade-off
- Consistency vs latency trade-off
- Failure handling (fail open vs closed)
- Be ready for: per-endpoint limits, tiered access, real-time config updates

Verbal Walkthrough Practice

Set **12-minute** timer. Talk through entire design out loud. Record yourself to identify unclear sections.