

Rate Limiter - Discussion Topics

Architecture & Design

1. When would you choose token bucket vs sliding window log?

- Memory efficiency considerations
- Burst handling requirements
- Precision vs performance trade-offs

2. How do you implement rate limiting in a distributed system?

- Centralized vs distributed counters
- Redis-based implementations
- Eventual consistency challenges

3. Where should rate limiting be implemented in the architecture?

- API Gateway level
- Application level
- Database level
- Trade-offs of each approach

Real-World Scenarios

1. Design a rate limiter for a public API with multiple pricing tiers

- Different limits per tier
- Graceful degradation
- Quota management

2. How would you handle rate limiting for a viral event?

- Sudden traffic spikes
- Fair distribution of resources
- Protecting critical services

3. Design rate limiting for a payment processing system

- Fraud prevention
- User experience balance
- Compliance requirements

Performance & Optimization

1. How do you minimize latency impact of rate limiting?

- Local caching strategies
- Async counter updates
- Optimistic vs pessimistic approaches

2. What metrics would you monitor for rate limiting effectiveness?

- Rejection rates
- Latency percentiles
- False positive rates

Advanced Topics

1. How do you handle rate limiting across multiple regions?

- Global vs regional limits
- Synchronization strategies
- Consistency models

2. What's your approach to adaptive rate limiting?

- Dynamic limit adjustment
- Load-based throttling

- Machine learning approaches

3. How do you prevent rate limit bypass attempts?

- IP rotation detection
- Fingerprinting techniques
- Behavioral analysis