

Epic Custom Load Balancer

Kasiakou M.M.

1. Target Audience & Key Metrics

1.1 Target Audience:

- **DevOps Engineers:** Manage deployment, scaling, and monitoring.
- **Site Reliability Engineers (SRE):** Ensure high availability and performance.
- **Backend Developers:** Integrate services with the load balancer.

1.2 Key Metrics

Metric	Description
DAU (Daily Active Users)	Measures daily traffic to estimate load.
RPS (Requests Per Second)	Critical for scaling backend capacity.
Error Rate	Tracks failed requests (5xx/4xx errors).
Uptime	Target ≥99.9% availability (SLAs).
Retention	Ensures consistent performance for returning users.

2. Technical Requirements

2.1 Performance & Scalability

- **RPS Handling:** Minimum **500+ RPS** with low latency (<100ms).
- **Algorithm Support:** Round Robin, Least Connections, and may be etc.

2.2 Storage & Configuration

- **Database:** Stores server lists, routing rules, and health check logs.

2.3 Architecture Components

1. Load Balancer Core

- Request routing based on selected algorithm.
- Health checks.
- Dynamic configuration (DB-driven updates).

2. Backend Pool

- Scalable microservices/containers/VMs.
- Automatic failover (unhealthy node removal).

3. System Architecture Diagram

