

Load Balancing

Computing

How traffic is spread across servers, and what each algorithm assumes.

The common algorithms and where each fits.

Algorithm	Sends to	Best when
Round robin	Each server in turn	Requests cost the same
Least connections	Fewest active connections	Requests vary in duration
Weighted	Proportional to capacity	Servers differ in size
IP hash	Same client → same server	Sticky sessions needed
Least response time	Fastest responder	Mixed server health

Health checks matter more than the algorithm. A balancer that keeps sending traffic to a dead node is worse than none.