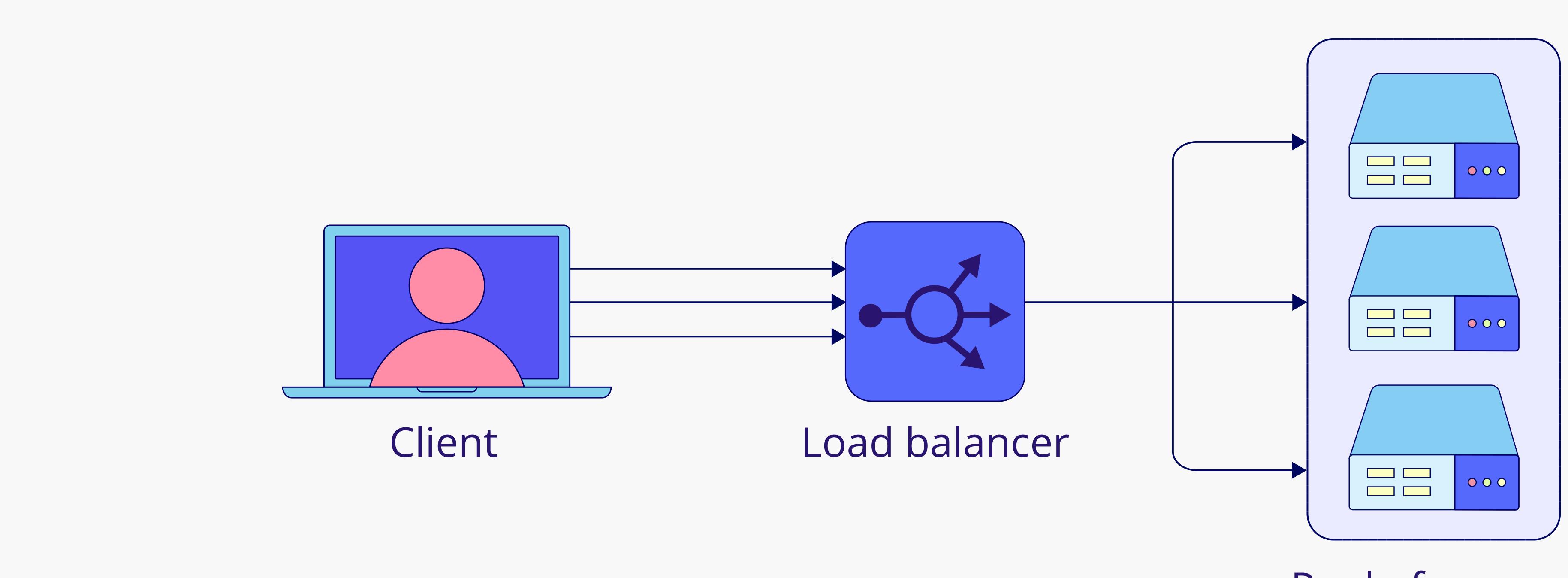


Load Balancing

Load balancing is the uniform distribution of incoming network traffic across multiple servers.



Pool of servers

Advantages:

Scalability

Adds/Removes servers to handle varying loads.

Reliability

Prevents failures by redirecting traffic to other functioning servers if one server fails.

Security

Helps mitigate DDoS attacks.

Performance

Routes request to less loaded servers.

Disadvantages:

Single point of failure

If one load balancer is used, its failure can bring down the entire system.

Cost

There is a cost for additional hardware and/or software.

Latency

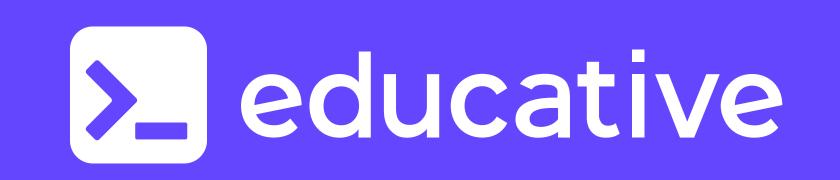
The load balancer's extra layer of routing requests can introduce a slight delay.

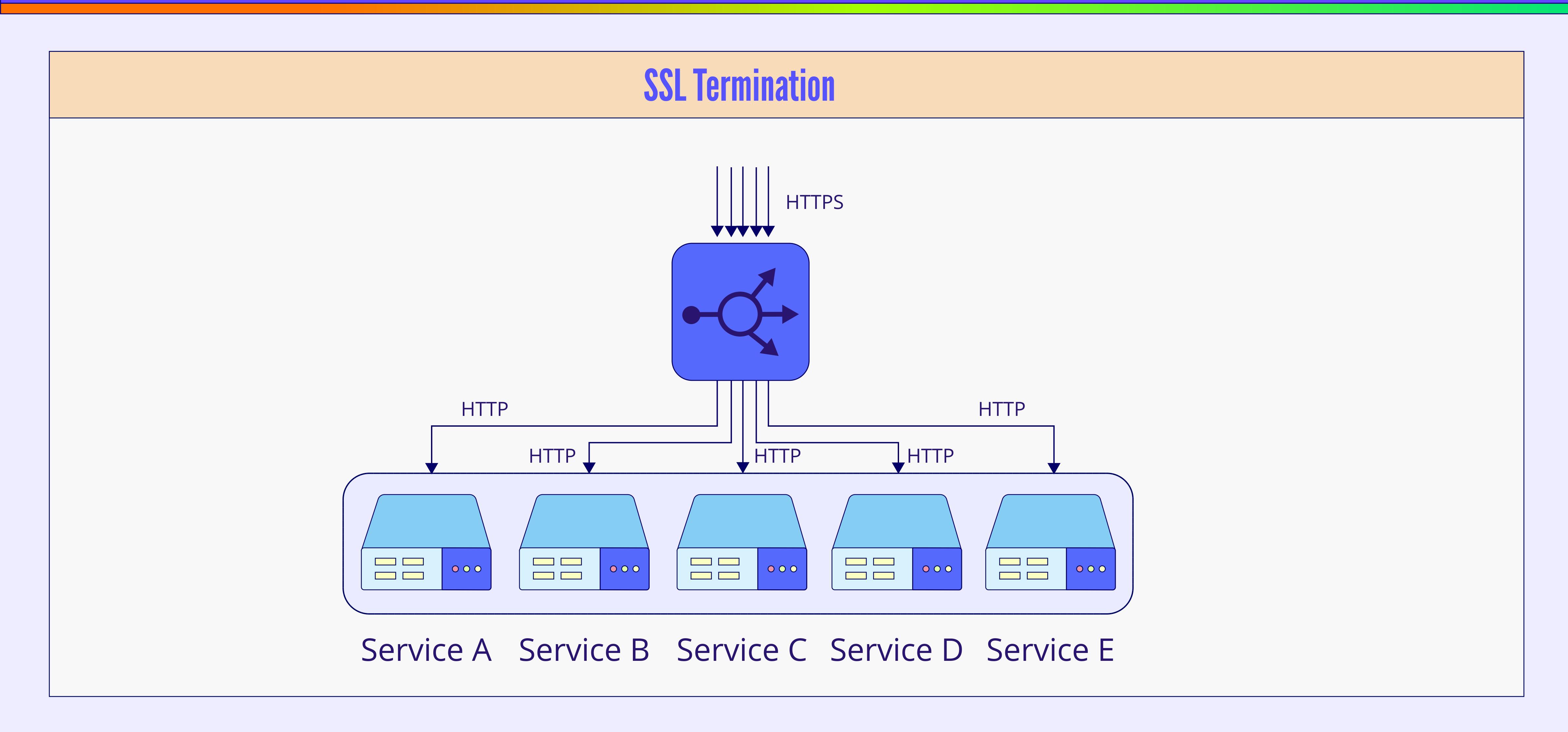
Complexity

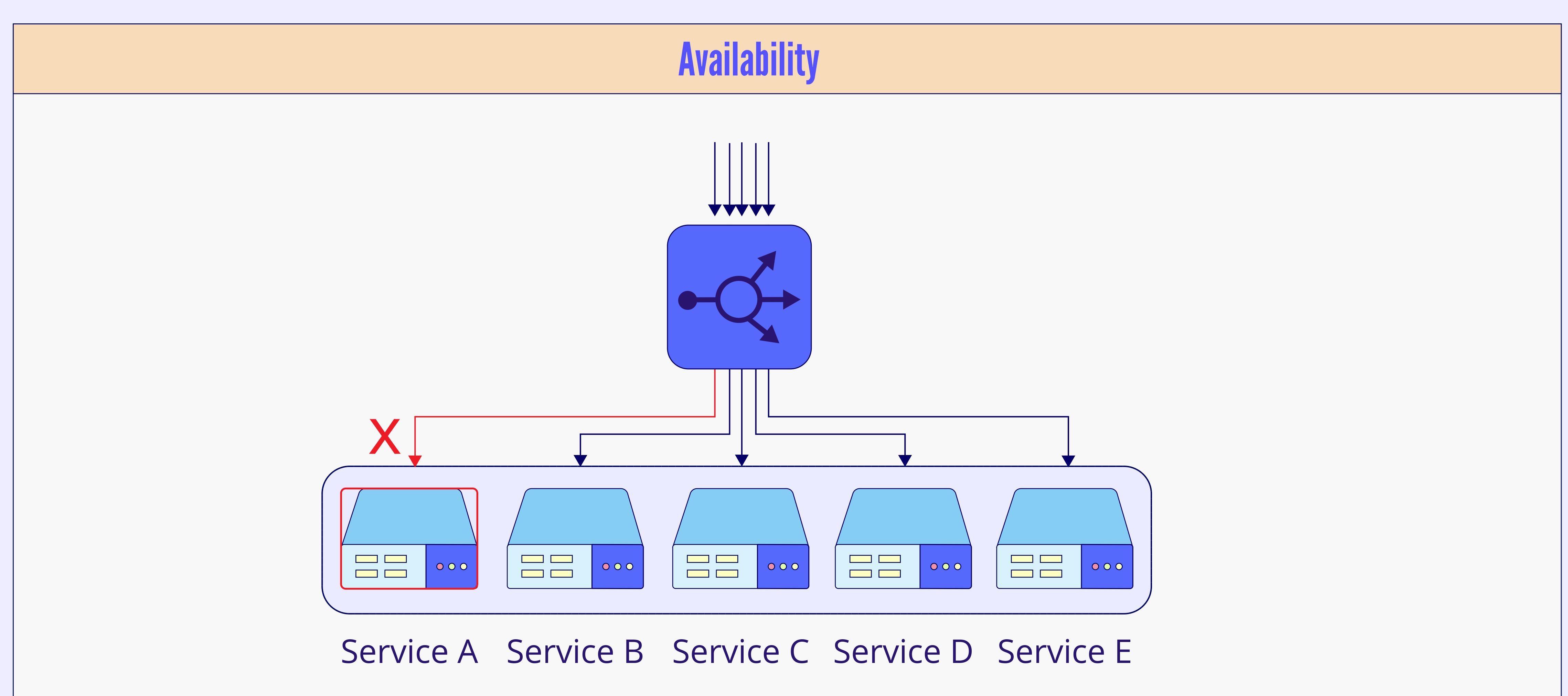
Requires specialized knowledge for configuration and ongoing management.

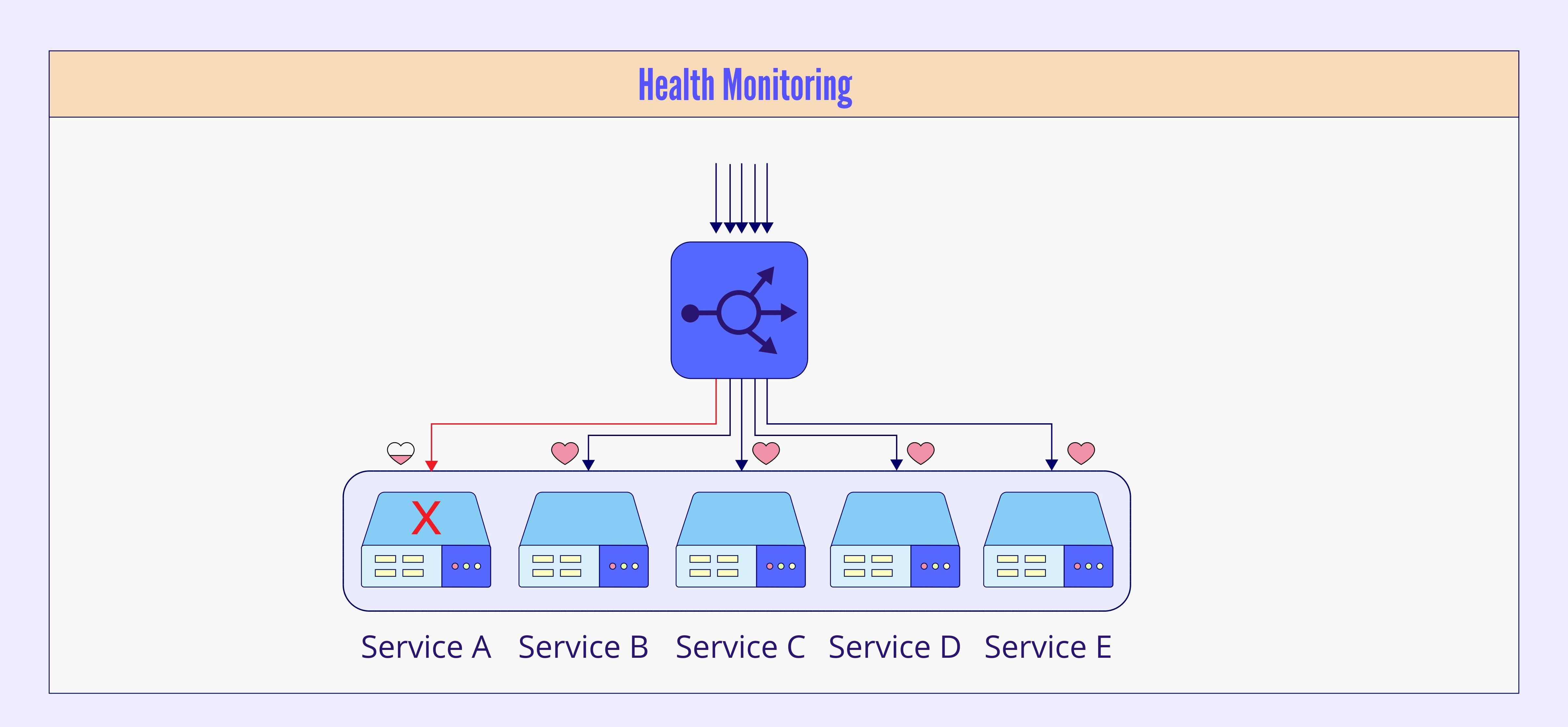
Use cases

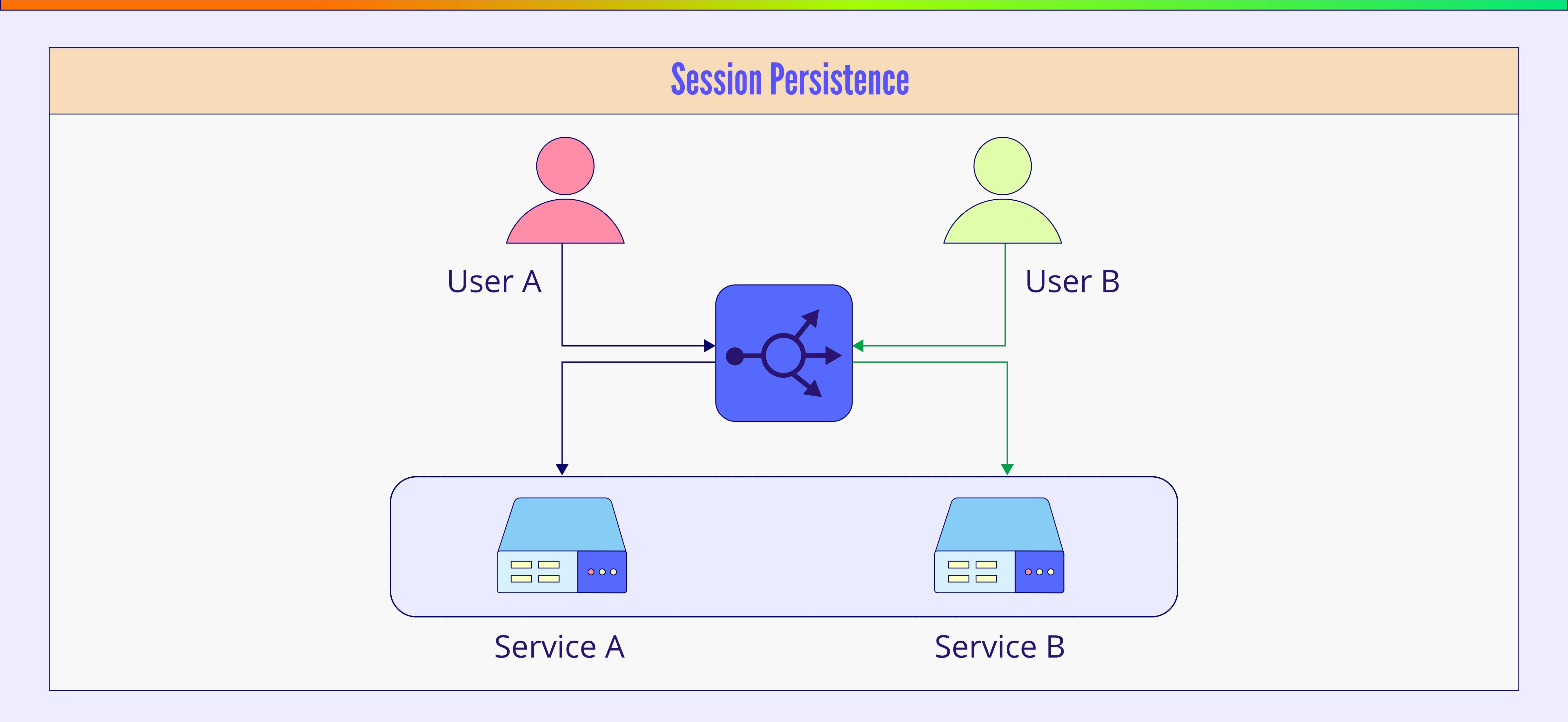
Service A Service C Service D Service E





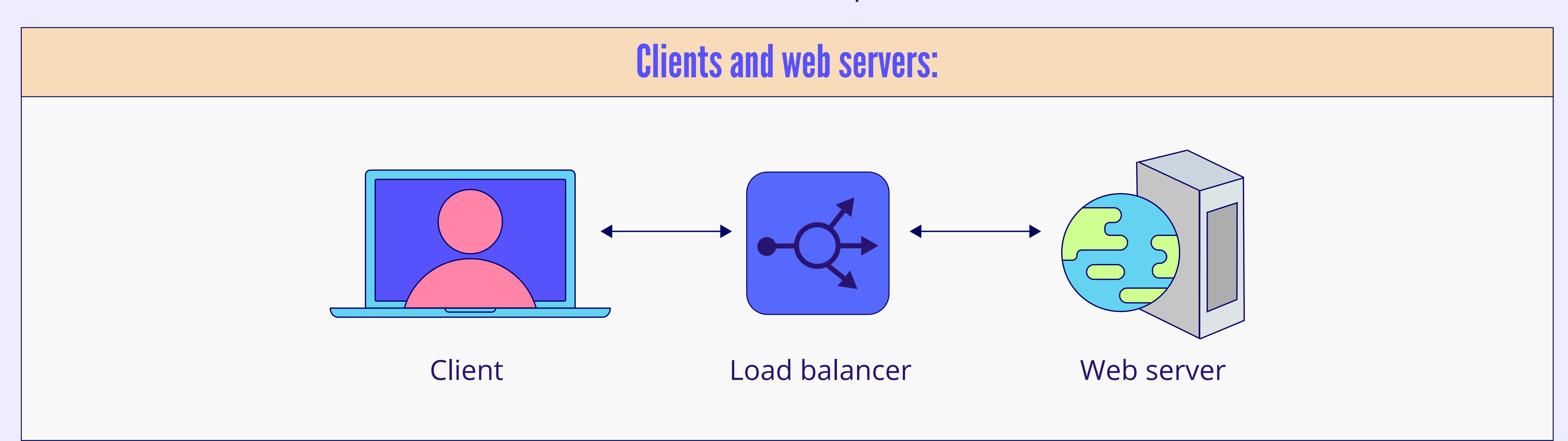


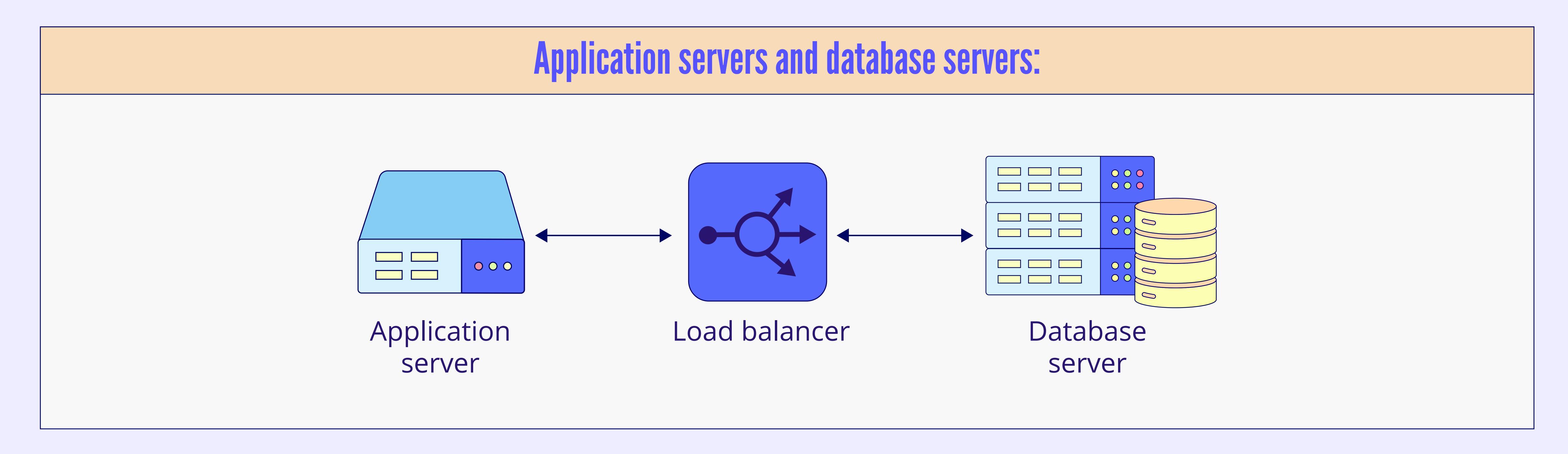


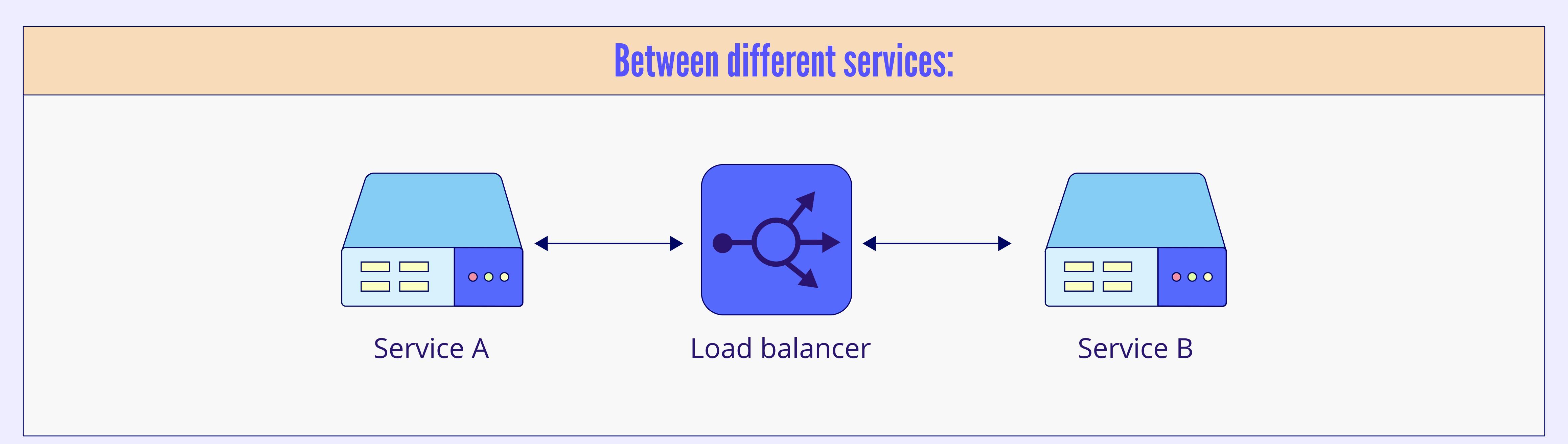


Where to Place Load Balancer

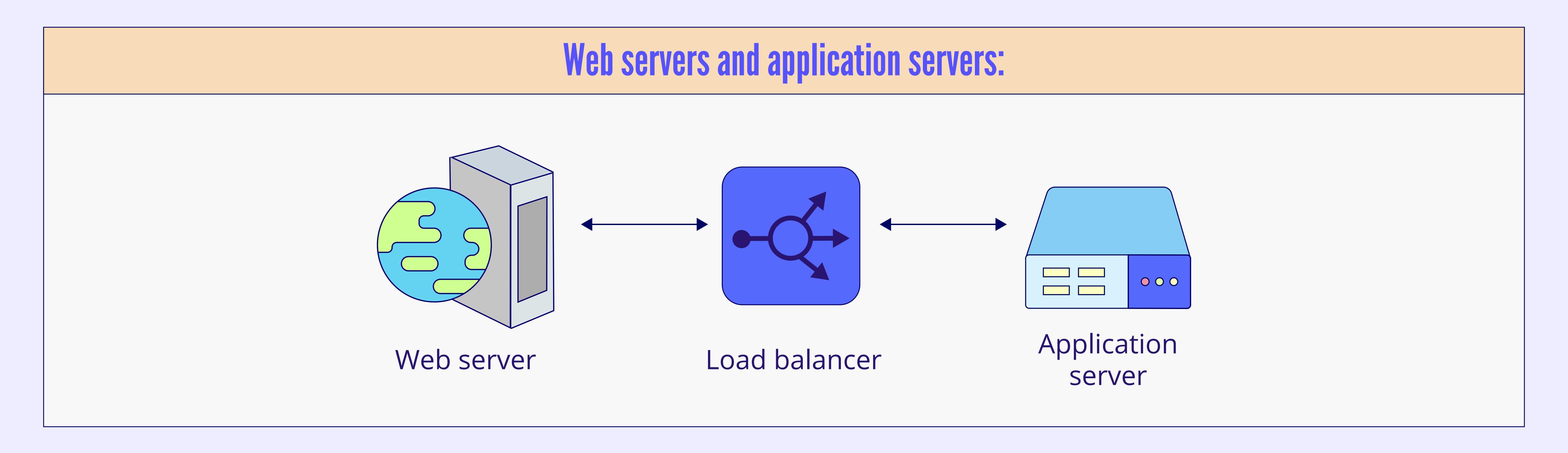
The load balancer can be placed between:



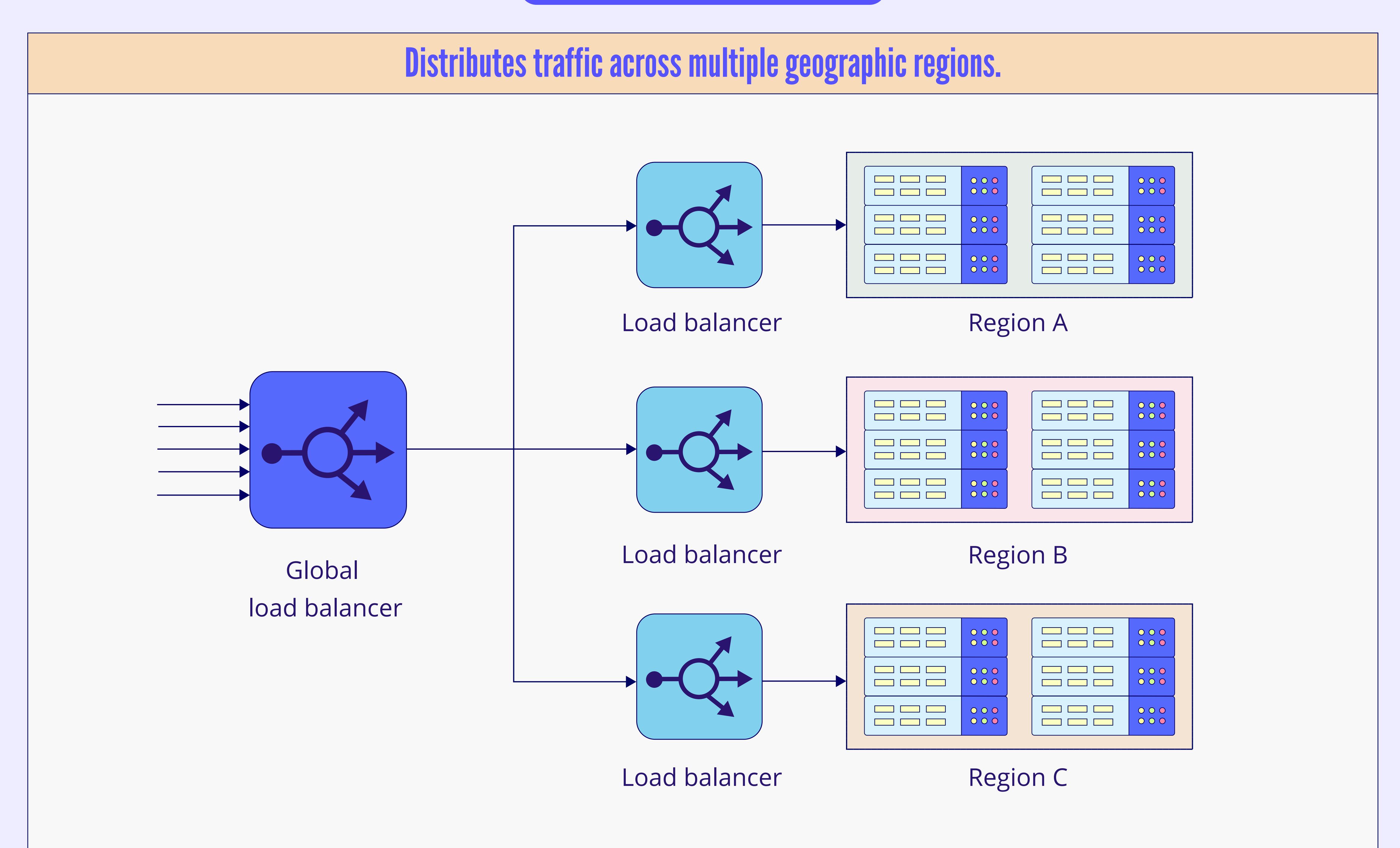








Global Load Balancing



Local Load Balancing

