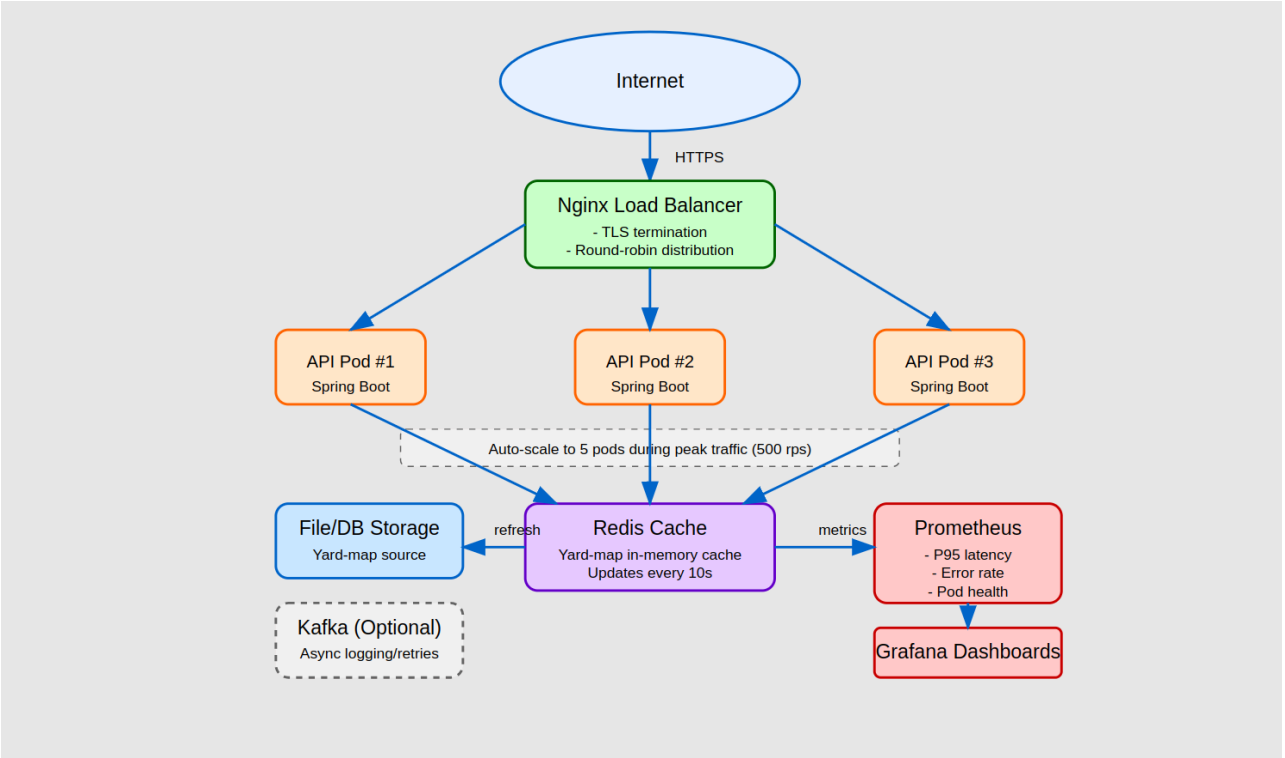
PickSpot API System Design

PHILIP SIMON DEROCK
philipsimonderock@gmail.com
8300057632



The pickSpot API system architecture divides traffic among three to five Spring Boot API pods (scaling according to demand) using a Nginx load balancer. These pods can handle 100 to 500 rps while keeping response times under 300 ms. For O(1) yard-map lookups, each pod uses a shared Redis cache, with data being refreshed from the source file or database every ten seconds. Resilience is ensured by the stateless design; in the event of a pod failure, traffic automatically switches to the remaining pods without losing data, and Redis failures cause a fallback to the most recent good data. When SLOs are violated, Prometheus and Grafana send out alerts based on their monitoring of P95 latency, error rates, and pod health. Consistent performance is guaranteed even during traffic spikes and updates thanks to the zero-downtime blue-green deployment process and horizontal auto-scaling rules (adding pods at >70% CPU, removing at <30% CPU).