

Aufgabe 6B: Load Balancing Theorie

Aufgabe 6B: Load Balancing Theorie

1. **Bedeutung von Load Balancing:**

Load Balancing sorgt für eine gleichmässige Verteilung des Datenverkehrs auf mehrere Server. Dies verhindert Überlastungen einzelner Systeme und gewährleistet eine stabile sowie performante IT-Infrastruktur.

2. **Funktionsweise:**

Ein Load Balancer analysiert eingehende Anfragen und verteilt den Traffic anhand definierter Algorithmen (z. B. Round Robin) auf verschiedene Server. Somit trägt jeder Server einen angemessenen Anteil an der Last.

3. **Vorteile:**

Load Balancer verbessern die Servicequalität, da sie Lastspitzen abfangen, und erhöhen die Verfügbarkeit der Dienste. Fällt ein Server aus, wird der Traffic automatisch auf andere Systeme umgeleitet, wodurch Engpässe vermieden und die Performance optimiert wird.

4. **Round Robin Algorithmus:**

Beim Round Robin werden Anfragen in einer festgelegten Reihenfolge an alle verfügbaren Server verteilt. Jeder Server erhält nacheinander eine Anfrage, was effektiv ist, wenn die Ressourcennutzung der Anfragen vergleichbar ist.

5. **Optimierung von Leistung und Verfügbarkeit:**

Durch den Einsatz eines Load Balancing-Systems wird der Traffic gleichmässig verteilt. Dadurch wird nicht nur die Leistung und Skalierbarkeit der Dienste optimiert, sondern auch Redundanz aufgebaut, die den Betrieb auch bei Ausfällen einzelner Komponenten sichert.

6. **Beispielprodukt – HAProxy:**

HAProxy wird häufig zur effizienten Verteilung von HTTP-Traffic eingesetzt. Ein typisches Tutorial demonstriert die Konfiguration von HAProxy, inklusive Funktionen wie Health Checks und Session Persistence, um eine zuverlässige Lastverteilung zu erreichen.

7. **Voraussetzungen an die IT-Infrastruktur:**

Für den Einsatz von HAProxy sind ein dedizierter Server oder eine VM mit ausreichender Bandbreite und ein aktuelles Linux-Betriebssystem erforderlich. Zudem muss die Netzwerkinfrastruktur – inklusive Firewall und Routing – optimal konfiguriert sein.

8. **Einsatzszenarien:**

Load Balancing ist sinnvoll, sobald Server an ihre Leistungsgrenzen stoßen oder bei stetig wachsendem Traffic. Insbesondere bei hoher Verfügbarkeit und schnellen Reaktionszeiten ist ein Load Balancing-System essenziell.