

System Load

Die **System Load** (oder *Load Average*) eines Linux-Servers gibt an, wie stark die Systemressourcen – primär die CPU und I/O-Wartezeiten – über einen bestimmten Zeitraum ausgelastet sind. Sie misst die **durchschnittliche Anzahl von Prozessen, die sich im Zustand "ausführbar" (runnable) oder "nicht unterbrechbar" (uninterruptible) befinden**.

Die 3 Werte der Load Average

Wenn man Befehle wie `uptime`, `top` oder `w` ausführt, sieht man drei Zahlenwerte für die Load Average:

Plaintext

```
load average: 0.75, 1.20, 2.05
```

Diese drei Werte repräsentieren den geglätteten Durchschnitt der Load über drei verschiedene Zeitintervalle:

1. **1 Minute** (0.75)
2. **5 Minuten** (1.20)
3. **15 Minuten** (2.05)

Anhand der Entwicklung lässt sich erkennen, wohin die Tendenz geht:

- **Steigend** (z. B. $0.50 \rightarrow 1.20 \rightarrow 2.50$): Die Last nimmt aktuell zu.
- **Fallend** (z. B. $2.50 \rightarrow 1.20 \rightarrow 0.50$): Das System kühlt sich gerade ab.

Was bedeuten die Prozess-Zustände genau?

Linux zählt für die Berechnung der Load zwei Prozess-Zustände im Kernel:

1. **Running / Runnable (R)**: Prozesse, die entweder gerade auf einer CPU ausgeführt werden oder darauf warten, von der CPU Zeit zugewiesen zu bekommen (Prozessorschlange).
2. **Uninterruptible Sleep (D)**: Prozesse, die auf eine I/O-Ressource warten (z. B. Lesen/Schreiben auf der Festplatte/SSD, Dateizugriffe über NFS, Netzwerk-I/O). Während sie warten, können sie nicht unterbrochen werden.

“ **Wichtig:** Eine hohe Load bedeutet nicht zwangsläufig, dass die CPU zu 100 % ausgelastet ist. Eine langsame oder blockierte Festplatte kann viele Prozesse in den **D**-Zustand versetzen und die Load extrem in die Höhe treiben, obwohl die CPU fast nichts zu tun hat (*I/O Wait*).

Wie interpretiert man die Zahl? (CPU-Kerne berücksichtigen)

Eine nackte Load-Zahl lässt sich nur im Kontext der **Anzahl der verfügbaren CPU-Kerne (bzw. Threads)** bewerten.

- **1 CPU-Kern:**

- **Load = 1.00:** Der Kern ist zu genau 100 % ausgelastet. Jeder Prozess bekommt sofort Rechenzeit.
- **Load = 0.50:** Der Kern ist zu 50 % ausgelastet.
- **Load = 2.00:** Das System läuft über der Kapazität. Durchschnittlich wartet 1 Prozess darauf, verarbeitet zu werden (50 % Überlastung).

- **4 CPU-Kerne:**

- **Load = 4.00:** Alle 4 Kerne sind optimal zu 100 % ausgelastet.
- **Load = 2.00:** Die CPU-Kapazität ist etwa zu 50 % ausgelastet.
- **Load = 8.00:** Die CPU-Kapazität ist doppelt überlastet.

Zustand	1 CPU-Kern	4 CPU-Kerne	Interpretation
---------	------------	-------------	----------------

Unterauslastet	< 0.70	< 2.80	Das System hat reichlich Reserven.
Optimal	0.70 – 1.00	2.80 – 4.00	Hohe Effizienz ohne Wartezeiten/Staus.
Überlastet	> 1.00	> 4.00	Prozesse müssen warten; Reaktionszeiten steigen.

Die Anzahl der verfügbaren logischen Kerne lässt sich auf dem Server wie folgt abfragen:

Bash

```
nproc
```

Der Unterschied zwischen CPU-Auslastung (%) und Load

- **CPU-Auslastung (in %):** Misst rein die verbrauchte Rechenzeit der CPU in einem bestimmten Moment. Sie kann maximal 100 % pro Kern erreichen.
- **System Load:** Ist ein kontinuierlicher Indikator, der auch **Warteschlangen** abbildet. Sie ist nach oben offen (eine Load von 50.0 auf einem System mit 4 Kernen zeigt eine extreme Überlastung an).

Diagnose-Befehle zur Überprüfung

Um herauszufinden, warum die Load hoch ist, bieten sich folgende Terminal-Werkzeuge an:

1. `uptime` / `w`: Schneller Überblick über die 1-, 5- und 15-Minuten-Load.
2. `top` / `htop`: Zeigt Echtzeit-Prozesse, CPU-Auslastung sowie den Zustand `%wa` (I/O-Wait).
3. `dstat` / `iostat`: Hilft festzustellen, ob eine Festplatte oder SSD der Flaschenhals ist.