



Wie effizient sind Rechenzentren wirklich?

Umfassende Bestimmung der Energieeffizienz | Der Markt für Rechenzentren ist in Bewegung - genauso wie der entsprechende Stromverbrauch. Bisher wurde die Energieeffizienz oft mit dem PUE ausgedrückt, dieser wird aber der Realität immer weniger gerecht. Wie lassen sich Effizienz und Emissionen von Rechenzentren genau bestimmen?

BABAK FALSAFI

Die Strommärkte waren in den letzten zwei Jahren aufgrund von Bedenken bezüglich der Energieversorgung äusserst volatil. Mit der zunehmenden Digitalisierung und der Einführung von Künstlicher Intelligenz in allen Branchen wächst nicht nur die Besorgnis hinsichtlich des Stromverbrauchs in Rechenzentren, sondern auch in Bezug auf den steigenden Bedarf und die ökologischen Auswirkungen der IT. Eine Studie von Schneider Electric prognostiziert für den gesamten IT-Sektor zwischen 2023 und 2030 einen jährlichen Anstieg des Stromverbrauchs um 5%,

wobei 75% davon auf Rechenzentren (getrieben durch KI) und Mobilfunknetze (aufgrund des Übergangs zu 5G) entfallen sollen.

Der Stromverbrauch in Rechenzentren weltweit ist in den letzten zehn Jahren relativ stabil geblieben und macht etwa 1,5% des gesamten Strombedarfs aus. In dienstleistungsorientierten Wirtschaftsräumen ist dieser Anteil etwas höher, beispielsweise beträgt er in der Schweiz 4%. Diese Stabilität ist darauf zurückzuführen, dass Unternehmen vermehrt in die Cloud umziehen und Cloud-Anbieter den Stromverbrauch im Auge behalten,

um ihre Investitionsrendite zu maximieren. Co-Locator-Rechenzentren, in denen die IT-Ausrüstung der Kunden untergebracht ist, entwickeln kontinuierlich effizientere Infrastrukturen für Kühlung, Stromverteilung und Wärmerückgewinnung.

Eine Studie des Uptime Institute (**Bild 1**) aus dem Jahr 2022 zeigt, dass die Energieeffizienz von Rechenzentren weltweit in den letzten Jahren nur langsam abnimmt - gemessen an der konventionellen Kennzahl Power Usage Efficiency (PUE), die den Anteil der in der Gebäudeinfrastruktur verbrauchten Elektrizität angibt.

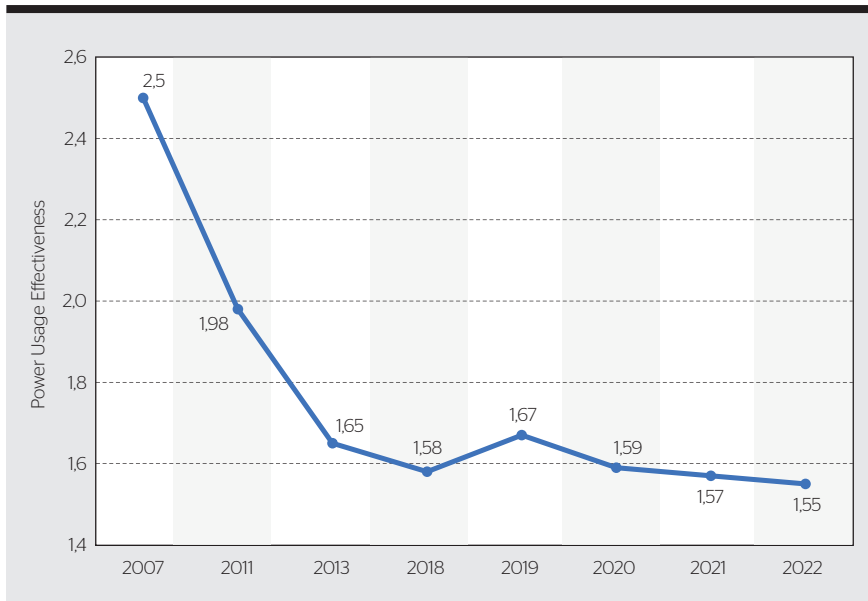


Bild 1 Historische Entwicklung des PUE.

Was sind die Grenzen von PUE?

Ähnlich wie viele andere Kennzahlen hat auch der PUE seine Grenzen. Zunächst einmal berücksichtigt er nicht die gesamten Umweltauswirkungen des Rechenzentrumsbetriebs. Er erfasst lediglich den Anteil des Stroms in der Gebäudeinfrastruktur, einschliesslich Kühlung und Stromverteilung, und vernachlässigt verschiedene Möglichkeiten, wie der gesamte Energiefluss im Rechenzentrum zur Nachhaltigkeit beitragen oder Emissionen reduzieren kann. Moderne Rechenzentren setzen sowohl auf Technologien zur Wiederverwertung von Abwärme als auch auf erneuerbare Energien vor Ort.

Zweitens unterliegt der PUE-Wert Schwankungen, die von Faktoren wie der Jahreszeit, der aktuellen Auslastung des Rechenzentrums und sogar der Tageszeit abhängen. Dies macht ihn zu einem unzuverlässigen Massstab für die Effizienz, besonders für Rechenzentren, die unter wechselnden Umweltbedingungen und Auslastungen arbeiten.

Eine der wohl gravierendsten Einschränkungen des PUE-Wertes besteht darin, dass er bei der Messung der IT-Effizienz wenig aussagekräftig ist. Ironischerweise können nämlich ineffiziente Server den PUE-Wert erstaunlich niedrig erscheinen lassen. Dies liegt daran, dass sich der PUE-Wert verbessert, je mehr Energie von den IT-Geräten verbraucht wird. Dies

schafft einen Anreiz, überschüssige IT-Ressourcen bereitzustellen, um die PUE-Werte künstlich zu verbessern. Selbst wenn die IT-Ausstattung effizient ist, was in neu errichteten Rechenzentren wahrscheinlich der Fall ist, beeinflusst auch die Auslastung der IT-Systeme die betriebliche Effizienz erheblich – der PUE-Wert gibt jedoch keine Auskunft darüber, ob die Server zu 20 % oder 80 % ausgelastet sind.

In der Schweiz strebt man dank wegweisender Arbeit im Bereich nachhaltiger Rechenzentren einen PUE-Wert von 1,15 an. Das bedeutet, dass mehr als 80 % des Stroms in diesen Rechenzentren für die IT-Ausstattung (Server, Speicher, Netzwerk) genutzt werden.

Die Frage nach dem Stromverbrauch in Rechenzentren und seiner Effizienz dreht sich somit um die IT.

Wohin entwickelt sich die IT?

Die technologischen Prognosen deuten für den Stromverbrauch in der IT auf ein dramatisches Wachstum hin. Auf der einen Seite profitierten Silizium-Fertigungstechnologien vier Jahrzehnte lang von einer Verdoppelung der Chipdichte alle zwei Jahre (das sogenannte Mooresche Gesetz). Diese Steigerung der Chipdichte ging mit einer entsprechend verbesserten Energieeffizienz einher, sodass dichtere Chips mit höheren Frequenzen betrieben werden konnten, ohne dass der Gesamtstromverbrauch anstieg.

Aber die Fortschritte bei der Siliziumdichte sind inzwischen an physikalische Grenzen gestossen. Obwohl es Verbesserungen bei den Algorithmen, der Software und dem Chipdesign gibt, die eine Spezialisierung der Plattformen ermöglichen, wird keine von ihnen zu einer exponentiellen Steigerung der Dichte für alle Dienste in Rechenzentren führen.

Auf der anderen Seite ist die Nachfrage aufgrund des rasanten Wachstums der Künstlichen Intelligenz (KI) in den letzten zehn Jahren um das Sechsfache pro Jahr gestiegen. Dies geschieht in Verbindung mit der Verlangsamung des Mooreschen Gesetzes, was bedeutet, dass neue IT-Geräte nun schneller als zuvor entwickelt und bereitgestellt werden müssen.

Bild 2 veranschaulicht den rasanten Anstieg der Thermal Design Power (TDP), also der maximalen Wärme-

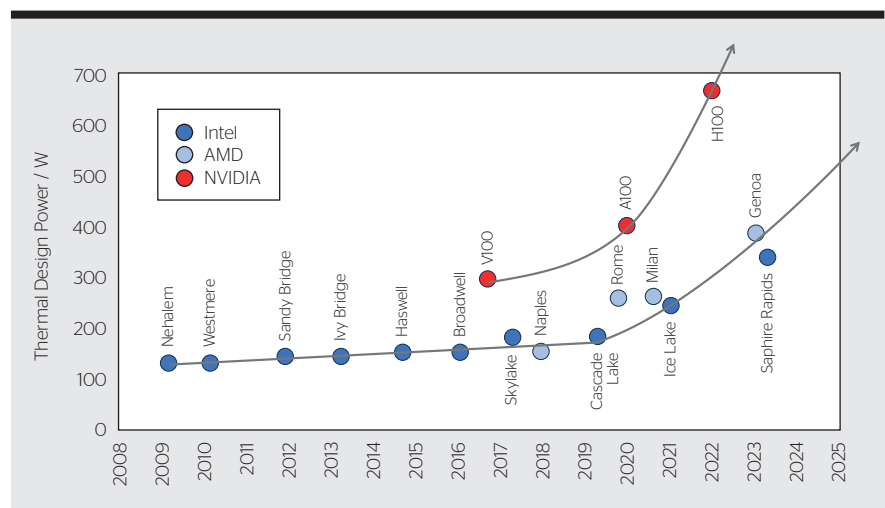


Bild 2 Anstieg der maximalen Wärmeabgabe in den letzten Jahren.

abgabe in Watt, von Server-CPUs, den grundlegenden Recheneinheiten in Rechenzentren. Die TDP von CPUs ist seit den 1990er-Jahren von einem einstelligen Wert auf etwa 100 W im Jahr 2000 angestiegen und hat sich dann dank energieeffizienter Designs für etwa zehn Jahre stabilisiert. Da die Effizienz- und Dichtegewinne immer kleiner werden, steigen die TDPs bei den neuesten CPUs jedoch rapide an. Grafikprozessoren, GPUs, die die Grundlage für Künstliche Intelligenz (KI) bilden, weisen einen dramatisch höheren TDP-Wert auf: Die Nvidia A100 hat 300 W und das neueste Produkt, das H100, sogar 700 W.

Diese Trends erfordern geeignete Kennzahlen und Methoden zur Bewertung der Energieeffizienz von IT-Geräten und -Lasten.

Was sind die richtigen Messgrößen für die IT?

Angesichts der Wachstumsprognosen für den IT-Stromverbrauch und der Fortschritte in der Infrastruktur von Rechenzentren sind neue Kennzahlen und Methoden zur Bewertung der Energieeffizienz und der Emissionen in Rechenzentren erforderlich. Dies gilt sowohl für die Infrastruktur als auch für die IT-Ausrüstung. Diese Kennzahlen müssen nicht nur die Wiederverwertung von Wärme und den Einsatz erneuerbarer Energien in der Gebäudeinfrastruktur berücksichtigen, sondern auch die Effizienz verschiedener Komponenten der IT-Ausrüstung, einschliesslich Rechenlogik (z. B. CPUs, GPUs und Beschleuniger), Speicher, Datenspeicherung und Netzwerkausrüstung. Darüber hinaus sind präzise Messmethoden sowie geeignete Software- und Hardwareinstrumente nötig, um diese Kennzahlen zu ermitteln.

Es ist ebenso entscheidend, sich mit anderen Kennzahlen auseinanderzusetzen, die bisher eher im Hintergrund standen, aber zunehmend an Bedeutung gewinnen. Die Auslastung der Workloads bietet beispielsweise einen differenzierteren Einblick in die Effizienz der Nutzung von Computerressourcen. Ein Server, der die meiste Zeit nur zu 20% ausgelastet ist, stellt nicht nur eine ungenutzte Ressource dar, sondern auch eine eklatante Ineffizienz, die direkt zu Energieverschwendung und höheren Betriebskosten führt.

Zentral ist auch die technologische Qualität. Obwohl dies keine herkömmliche Kennzahl ist, fungiert sie als Massstab zur Bewertung der Geräte und Methoden, die in einem Rechenzentrum verwendet werden. Betreiber sollten bei der Auswahl von Technologien auf Höchstleistung achten und sich auf die modernsten und effizientesten Optionen konzentrieren. Zum Beispiel kann die Entscheidung für Flash-Speicher anstelle von Festplattenlaufwerken den Stromverbrauch und den Kühlungsbedarf drastisch reduzieren und gleichzeitig den Datenzugriff beschleunigen. Ebenso steigert die Wahl von Glasfaserkabeln anstelle von Kupferkabeln für Netzwerke nicht nur die Geschwindigkeit, sondern minimiert auch den Stromverbrauch. Diese Auswahl erstreckt sich auch auf Server, Netzteile und Stromverteiler, die sich durch Energieeffizienz, Zuverlässigkeit und Langlebigkeit auszeichnen und somit sowohl den Energieverbrauch als auch die Gesamtbetriebskosten senken.

Und schliesslich ist die zulässige maximale Betriebstemperatur eine weitere wichtige Kennzahl. Traditionell wurden Rechenzentren bei niedrigeren Temperaturen betrieben, um das Risiko einer Überhitzung zu minimie-

ren. Moderne Geräte wurden jedoch für den sicheren Betrieb bei höheren Temperaturen entwickelt. Durch die Anpassung der maximal zulässigen Betriebstemperaturen an diese höheren Grenzwerte können Unternehmen den Energiebedarf für die Kühlung drastisch senken, was sich erheblich auf die Gesamteffizienz des Rechenzentrums auswirkt.

Das SDEA-Label ist umfassend

Um die Effizienz und die Emissionen des Rechenzentrumsbetriebs ganzheitlich messen zu können, wurde 2020 das SDEA-Label ins Leben gerufen; von der Swiss Datacenter Efficiency Association (SDEA) – einem Konsortium aus Nachhaltigkeitspionieren aus Industrie und Wissenschaft. Das KPI-Tool von SDEA bietet einen Rechner, der das Wärmerecycling, die Nutzung erneuerbarer Energien, die Konsolidierung und Virtualisierung von Arbeitslasten, die Nutzung von Server-, Speicher- und Netzwerkkomponenten, die Datenkomprimierung, erstklassige Komponententechnologie und die zulässige Betriebstemperatur erfasst. Gemäss einer aktuellen Studie der Internationalen Energieagentur IEA ist das SDEA-Label die einzige Zertifizierung für Rechenzentren, die ein quantitatives Ranking bietet, und nicht nur Empfehlungen ausspricht.

Link

→ www.sdea.ch



Autor

Babak Falsafi ist Professor an der Fakultät für Computer- und Kommunikationswissenschaften der EPFL und Präsident der SDEA.

→ EPFL, 1015 Lausanne
→ babak.falsafi@epfl.ch

Als gemeinnützige Organisation wird die SDEA teilweise durch das EnergieSchweiz-Programm des Bundesamts für Energie finanziert und kann Schneider Electric und SICPA als Premium-Sponsoren vorweisen.