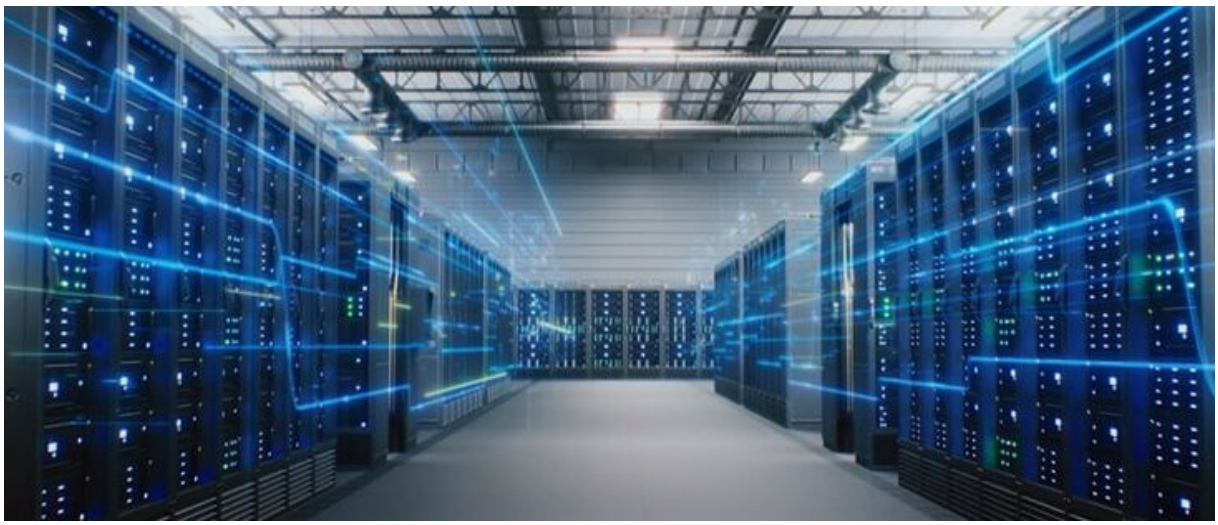
800 Volt statt 230: Warum KI-Rechenzentren ihre Energiearchitektur neu erfinden

19.06.2026, 10:27 Uhr
Kommentare: 1
Sicher arbeiten



KI treibt Rechenzentren an ihre Grenzen. © EvgeniyShkolenko/iStock/Getty Images Plus

KI treibt Rechenzentren an ihre Grenzen: Racks im Megawattbereich erzwingen neue Spannungsebenen und beschleunigen den Wechsel zu 800-Volt-Gleichstrom. Damit geraten klassische Versorgungskonzepte unter Druck – und Stromversorgung, Schutztechnik sowie Kühlung müssen als integriertes Gesamtsystem neu gedacht werden.

Leistungsdichte als treibender Faktor

Aktuelle KI-Beschleuniger nähern sich einer Leistungsaufnahme von 2.000 Watt je Chip, frühere GPU-Generationen lagen bei rund 500 Watt. Die Nvidia-Plattform Vera Rubin NVL144 soll ab der zweiten Jahreshälfte 2026 rund 120 bis 130 Kilowatt je Rack erreichen, die Kyber-Racks der Folgegeneration sind für 2027 mit 600 Kilowatt angekündigt.

Ein einzelnes Rack dieser Klasse zieht so viel Strom, wie rund 400 Haushalte verbrauchen. Die zugeführte Energie wandelt sich nahezu vollständig in Wärme um, womit Stromversorgung und Wärmeabfuhr eine gemeinsame Auslegung verlangen. In Deutschland sollte die Anschlussleistung der Rechenzentren laut Bitkom und Borderstep Institut 2025 auf 2.980 Megawatt steigen und Anfang 2026 die Marke von 3.000 Megawatt überschreiten, bis 2030 sind 5.000 Megawatt projektiert (www.bitkom.org/). Die KI-spezifische Leistung wächst dabei von 530 auf 2.020 Megawatt. Den Jahresstrombedarf der Rechenzentren für 2030 beziffert eine Borderstep-Studie auf 25 bis 35 Terawattstunden.

Grenzen der klassischen Versorgungskette

Die herkömmliche Versorgung führt über mehrere Stufen. Aus der Mittelspannung von zehn oder 20 Kilovolt setzt ein Transformator auf 0,4 Kilovolt herab, danach folgen unterbrechungsfreie Stromversorgung, eine 230-Volt-Wechselstromverteilung und serverinterne Wandler, die auf zwölf, fünf und 3,3 Volt heruntersetzen, bevor die Endregelung am Grafikprozessor bei rund einem Volt arbeitet. Jede Umwandlungsstufe kostet Wirkungsgrad und Bauraum. Bei den hohen Strömen aktueller Racks wachsen die ohmschen Verluste mit dem Quadrat der Stromstärke, sodass Leiterquerschnitte und Kupfermengen überproportional zunehmen. Eine rackinterne 54-Volt-Verteilung für ein Rack mit einem Megawatt erfordert Kupferschienen von bis zu 200 Kilogramm (www.datacenterdynamics.com/). Klassische Wechselstromkonzepte stoßen an dieser Stelle an eine physikalische Grenze.

Tipp der Redaktion



Das Nachschlagewerk für die Elektrofachkraft

Organisation, Durchführung und Dokumentation elektrotechnischer Prüfungen – „Elektrosicherheit in der Praxis“ unterstützt Sie bei der Umsetzung der Elektrosicherheit in Ihrem Unternehmen.

Jetzt testen!

Wechsel auf 800-Volt-Gleichstrom

Die Antwort der Branche heißt Hochspannungsgleichstrom mit 800 Volt, der zentral aus der Mittelspannung über Festkörpertransformatoren erzeugt und bis zum Rack geführt wird. Im Rack genügt dann eine einzige Gleichspannungswandlung, die rackinterne Wechselstromwandlung fällt weg. Gegenüber einer 480-Volt-Drehstromverteilung überträgt diese Architektur über 150 Prozent mehr Leistung durch denselben Kupferquerschnitt und macht die schweren Kupferschienen je Rack überflüssig. Der Gesamtwirkungsgrad von der Einspeisung bis zum Chip steigt von rund 83 auf über 92 Prozent, was einem Gewinn von bis zu fünf Prozent entspricht (https://developer.nvidia.com/). Nvidia und Infineon entwickeln diese Versorgung vom Netz bis zum Chip gemeinsam, die Halbleiterbasis bilden Silizium, Siliziumkarbid und Galliumnitrid. Als Zwischenschritt verfolgen mehrere Betreiber die 400-Volt-Gleichstromarchitektur des Open-Compute-Project-Vorhabens Mount Diablo. Der Serieneinsatz der vollwertigen 800-Volt-Technik ist mit der Rubin-Ultra-Plattform für 2027

angekündigt.

Lastdynamik und Spannungsqualität

KI-Lasten verhalten sich hochdynamisch. Das Training großer Modelle erzeugt eine anhaltende Grundlast im Megawattbereich, die Inferenz dagegen kurze Spitzen im Sekunden- bis Minutentakt, bei denen die Leistung in Größenordnungen von 100 Kilowatt in zehn Sekunden springen kann. Grafik- und Tensorprozessoren ziehen nichtlineare Ströme und speisen hohe Oberschwingungen samt Zwischenharmonischen ins Netz zurück. Rasche Lastsprünge führen zu Spannungseinbrüchen, Transienten und Flickern, einzelne Komponenten regeln unter Umständen nicht schnell genug nach. Greifen viele Anfragen gleichzeitig zu, addieren sich kurzzeitige Spitzen, die bis in den Bereich von zwei Gigawatt reichen können. Für die ausführende Elektrofachkraft rückt damit die Spannungsqualität in den Vordergrund. Eine fortlaufende Erfassung der Energiedaten nach DIN EN 50160 „Merkmale der Spannung in öffentlichen Elektrizitätsversorgungsnetzen“ und eine darauf abgestimmte Auslegung der unterbrechungsfreien Stromversorgung sichern den Betrieb unter diesen Lastprofilen ab.

Schutz- und Brandschutzkonzepte neu bewerten

Die veränderten Lastprofile verlangen eine Neubewertung der Schutztechnik. Leistungsschalter und Schutzkonzepte sind unter den steilen Lastgradienten und den hohen Bemessungsströmen anders auszulegen als bei gleichmäßiger Last, auch im Hinblick auf Selektivität und Störlichtbogenschutz. Hinzu kommt der Brandschutz, denn die Energiedichte je Rack und der Einsatz von Lithium-Ionen-Speichern in der Pufferung erhöhen das Risiko. Vorkonfigurierte KI-Container und Schaltschranklösungen halten die Brandschutzanforderungen nach DIN EN 1047-2 „Wertbehältnisse – Klassifizierung und Methoden zur Prüfung des Widerstandes gegen Brand – Teil 2: Datensicherungsräume und Datensicherungscontainer“ für Datensicherungsräume und DIN EN 13501 „Klassifizierung von Bauprodukten und Bauarten zu ihrem Brandverhalten – Teil 1: Klassifizierung mit den Ergebnissen aus den Prüfungen zum Brandverhalten von Bauprodukten“ für das Brandverhalten von Baustoffen ein. Brandfrüherkennung und Löschsysteme auf Basis inerter Gase schützen die Hardware, ohne den Betrieb durch Löschmittelschäden zu gefährden.

Downloadtipps der Redaktion

Unterweisung: Elektrofachkraft/verantwortliche Elektrofachkraft

Hier gelangen Sie zum Download.

Checkliste: Sichere Kabel- und Leitungsanlagen

Hier gelangen Sie zum Download.

Checkliste: Energiemanagmentsystem nach DIN VDE 0100-801

Hier gelangen Sie zum Download.

Kühlung als Teil der Energiearchitektur

Luftkühlung stößt bei diesen Leistungsdichten an physikalische Grenzen, weshalb die Wärmeabfuhr direkt am Chip erfolgt. Für rund ein Megawatt Kühlleistung sind ungefähr 1.500 Liter Wasser pro Minute nötig, die mehrere parallel arbeitende Pumpen in redundanter Anordnung bereitstellen. Direkte Flüssigkeitskühlung über Kühlplatten und Immersionskühlung lösen die Luftkühlung bei hoher Wärmelast ab. Beide Verfahren verknüpfen sich eng mit der Gleichstromversorgung, denn hohe Leistungsdichte und kontinuierliche Wärmeentwicklung lassen sich nur gemeinsam beherrschen. Die Wahl des Kühlmediums bringt einen Zielkonflikt mit sich, denn wassergestützte Systeme senken den Strombedarf, erhöhen aber den Wasserverbrauch. Bei dielektrischen Flüssigkeiten für die Immersionskühlung kann eine Novellierung der PFAS-Vorschriften (PFAS = per- und polyfluorierte Alkylsubstanzen) die Verfügbarkeit einschränken. Rücklauftemperaturen von bis zu 60 Grad Celsius erlauben es, die Abwärme in Fernwärmenetze einzuspeisen – ein Ansatz, den das Open Compute Project standardisiert.

Effizientere Chips senken die Last

Neben der Versorgungs- und Kühlarchitektur setzt die Forschung an der Recheneinheit selbst an. Frei programmierbare Logikbausteine, sogenannte FPGAs, erreichen bei der Inferenz tiefer neuronaler Netze je nach Modell und Stapelgröße Einsparungen von bis zu 90 Prozent gegenüber Grafikprozessoren – erzielt über Modellverschlankung und die Verteilung auf mehrere Bausteine. Ein weiterer Hebel sitzt im Datentransport, der in KI-Prozessoren 60 bis 90 Prozent des Stromverbrauchs ausmacht, und nicht in der Rechenoperation selbst. Ein an der Peking-Universität vorgestellter ferroelektrischer Transistor mit einer Gatelänge von einem Nanometer arbeitet bei 0,6 Volt und bringt Speicher und Logik auf nahezu gleiches Spannungsniveau, womit die Spannungswandler zwischen beiden Einheiten überflüssig werden. Solche Bausteine senken den Energiebedarf an der Quelle, befinden sich jedoch noch im Forschungsstadium.

Modularität als Antwort auf kurze Innovationszyklen

Die Innovationszyklen der Beschleuniger verkürzen sich auf zwölf bis 18 Monate, der Planungshorizont für die Infrastruktur sinkt von rund zehn auf zwei Jahre. Gebäude und Basisinfrastruktur halten 15 bis 20 Jahre, die KI-Racks tauschen die Betreiber alle zwei bis vier Jahre. Modulare Serverräume und vorkonfigurierte Container fangen diese Diskrepanz auf, da sie Strom, Kühlung und Rack als abgestimmtes Gesamtsystem liefern und sich an steigende Leistungen anpassen lassen. Ein mittelständischer Industriebetrieb nimmt auf diesem Weg eine KI-Infrastruktur binnen fünf Monaten in Betrieb, ohne das bestehende Rechenzentrum umzubauen. Offen bleibt die Standardisierung der Schnittstellen, denn Stromstecker, Wasserkupplungen und Steuerungssysteme verschiedener Hersteller passen nur teilweise zusammen. Das deutsche Energieeffizienzgesetz fordert zugleich niedrige PUE-Werte (PUE = Power Usage Effectiveness) und die Nutzung der Abwärme, was die Auslegung zusätzlich bindet.

Ausblick

Der Übergang zur Gleichstromarchitektur steht erst am Anfang. Der breite Bestand bleibt zunächst bei Drehstrom mit Rack-Dichten bis rund 200 Kilowatt, die vollwertige 800-Volt-Versorgung greift mit den Hochdichteplattformen ab 2027. Für die Auslegung von Serverräumen verschiebt sich das Gewicht von der reinen Rechenleistung zur

abgestimmten Kette aus Spannungsebene, Schutztechnik, Kühlung und Messtechnik. Wer diese Kette als zusammenhängendes System plant, hält die nächste Generation der Beschleuniger ohne tiefgreifenden Umbau in Betrieb.

Kurz & knapp

- KI-Racks erreichen Leistungen im Megawattbereich
- Klassische AC-Versorgung stößt an physikalische Grenzen
- 800-Volt-Gleichstrom reduziert Verluste und Kupferbedarf
- Spannungsqualität und Lastdynamik werden kritische Faktoren
- Schutztechnik und Brandschutz müssen neu ausgelegt werden
- Kühlung und Stromversorgung wachsen zum Gesamtsystem zusammen
- Der Engpass verschiebt sich von der Rechenleistung zur Energieversorgung

Weitere Beiträge zum Thema

Fehler in Stromverteilungen mit KI erkennen – was heute funktioniert

Sicherheitsanforderungen an die Stromversorgung von Rechenzentren unter neuer EU-Regulierung: Technische Auswirkungen im Detail

KI-basiertes Lastmanagement für USV-Systeme in Rechenzentren und Industrieanlagen (Teil 1)

KI-basiertes Lastmanagement für USV-Systeme in Rechenzentren und Industrieanlagen (Teil 2)

KI-basiertes Lastmanagement für USV-Systeme in Rechenzentren und Industrieanlagen (Teil 3)

Umgang mit und Wartung von Lithium-Ionen-Batteriesystemen in Rechenzentren

Fortschrittlicher Brandschutz für elektrische Installationen und Schaltschränke in Rechenzentren

Autor:

Thomas Joos

freiberuflicher Publizist



Thomas Joos ist freiberuflicher Publizist und veröffentlicht neben seinen Büchern auch Artikel für verschiedene Medien wie dpa, Computerwoche und C’t.

Seit seinem Studium der medizinischen Informatik berät er auch Unternehmen im Bereich IT, Security und Absicherung von Rechenzentren.