Nationale Rechenzentrumsstrategie

Vorschläge der Universität Bielefeld im Rahmen des Konsultationsprozess

0 - Abstract

Der Rechenzentrumsstandort Deutschland steht vor einem tiefgreifenden Wandel. Während 2023 laut Schneider Electric noch rund 95 % der KI-Workloads zentral verarbeitet wurden und nur 5 % in der Edge, wird bis 2028 ein Gleichgewicht von 50:50 erwartet – bei gleichzeitig stark wachsender Gesamtnachfrage auf 90 GW gegenüber 54 GW im Jahr 2023 [4]. Damit rückt insbesondere das **Edge-Computing in Kombination mit KI** in den kommenden fünf Jahren ins Zentrum der Herausforderungen und Chancen [1, 2, 3].

Ein zukunftsfähiger Standort Deutschland im Jahr 2030 muss sich durch **energie- und ressourceneffiziente Infrastrukturen, heterogene Hardwarelandschaften, offene Standards und resiliente Lieferketten** auszeichnen. Heterogene Hardware – also der gezielte Einsatz spezialisierter Beschleuniger (GPU, TPU, FPGA, ASIC, DPU) neben klassischen Prozessoren – bietet das Potenzial für **2–10x Effizienzgewinne bei zielgerichteten Workloads**, senkt Emissionen und stärkt technologische Souveränität. Voraussetzung dafür sind jedoch **interoperable, hardware-bewusste Software-Stacks, standardisierte Orchestrierung und qualifizierte Fachkräfte**, um die höhere Komplexität beherrschbar zu machen.

Deutschland verfügt über besondere Stärken im Bereich **Edge, IoT und Embedded-Systeme**, die gezielt in die Rechenzentrumsstrategie eingebettet werden sollten, um technologische Eigenständigkeit und Wettbewerbsfähigkeit auszubauen [1, 2].

Gleichzeitig bedarf es klarer politischer Rahmenbedingungen: **wettbewerbsfähige Energiepreise, Integration in erneuerbare Energiesysteme als intelligente Lasten, Förderung von Modularität und Wiederverwendbarkeit** sowie **innovationsfreundliche öffentliche Ausschreibungen mit stärkerem Zugang für KMU**.

Darüber hinaus kann Deutschland auf die Expertise aus **nationalen und europäischen Forschungs- und Entwicklungsprojekten** zurückgreifen. Die Arbeitsgruppe **Kognitronik und Sensorik** der Universität Bielefeld [12] verfügt über umfassende Erfahrung im Bereich heterogenes Rechnen für IoT, Edge, Cloud und HPC. Gemeinsam mit Industriepartnern - insbesondere **Christmann Informationstechnik** [13] - wurden modulare, heterogene Microserversysteme entwickelt, die auf offenen Computer-on-Module-Formfaktoren basieren. Projekte wie **M2DC** [5, 6], **VEDLIoT** [7, 8], **CAPE** [9, 10] oder **CoolEdgeHPC** [11] zeigen exemplarisch, wie **offene Architekturen, Modularität und energieeffiziente Kühlkonzepte** praxisnah umgesetzt und für die nächste Generation souveräner, nachhaltiger und leistungsfähiger Recheninfrastrukturen nutzbar gemacht werden können.

Das Zielbild 2030 ist ein **kognitives Compute-Continuum**, in dem Cloud- und Edge-Infrastrukturen nahtlos zusammenwirken, Rechenzentren energieproportional und nachhaltig betrieben werden, und offene, hardware-diverse Architekturen die Grundlage für Souveränität, Resilienz und Wettbewerbsfähigkeit bilden.

1 - Welche Merkmale und Rahmenbedingungen kennzeichnen aus Ihrer Sicht einen „zukunftsfähigen und leistungsstarken“ Rechenzentrumsstandort Deutschland im Jahr 2030?

Ein zukunftsfähiger und leistungsstarker Rechenzentrumsstandort Deutschland im Jahr 2030 zeichnet sich durch folgende Merkmale aus:

- **Kognitives Compute-Continuum**: Nahtlose Integration von Cloud- und Edge-Infrastrukturen, um Latenz, Effizienz und Verfügbarkeit für unterschiedliche Workloads sicherzustellen.
- **Energieintegration**: Enge Kopplung an erneuerbare Energien mit intelligenter Lastverschiebung und Rechenzentren als flexible Verbraucher im Energiesystem.
- **Nachhaltige Kühlung und Wärmenutzung**: Innovative Kühlkonzepte in Cloud und Edge, konsequente Abwärmenutzung sowie Gleichrangigkeit von PUE-Zielen in Edge- und Hyperscale-Umgebungen.
- **Heterogene Hardwarelandschaften**: Breiter Einsatz von Beschleunigern (GPU, TPU, FPGA, ASIC, DPU) und spezialisierten Kernen für KI/ML, Simulation, Kryptografie, Datenbanken und Netzinfrastruktur.
- **Stärkung des europäischen Edge-Ökosystems**: Aufbau einer eigenständigen Innovationslandschaft durch gezielten Einsatz spezialisierter, energieeffizienter Beschleuniger für Edge-Szenarien (z. B. TPU, FPGA, ASIC, DPU), die Workloads präzise abbilden und europäische Industrie- und Wissenschaftsanwendungen unterstützen.
- **Intelligente Software-Stacks**: Leistungsfähige, hardware-bewusste Software-Stacks und Orchestrierungswerkzeuge, die Workloads analysieren und optimal auf unterschiedliche Beschleuniger verteilen. Nur durch diese enge Verzahnung von Hard- und Software lassen sich die Potenziale heterogener Architekturen in der Praxis ausschöpfen.
- **Digitale Souveränität**: Förderung offener Standards und Architekturen (z. B. RISC-V), Aufbau vielfältiger Lieferketten zur Vermeidung von Abhängigkeiten.
- **Effizienz und Wirtschaftlichkeit**: 2–10x Effizienzgewinne bei zielgerichteten Workloads, höhere Ressourcenauslastung, reduzierte TCO sowie optimierte Wärmenutzung.

2 - Welche zentralen Herausforderungen und Chancen sehen Sie für den Rechenzentrumsstandort Deutschland in den kommenden Jahren?

Die zentralen Herausforderungen und Chancen für den Rechenzentrumsstandort Deutschland liegen in:

- **Integration von Cloud und Edge**: Die Gestaltung eines kohärenten „kognitiven Compute-Continuums“ ist essenziell, um Effizienz, Skalierbarkeit und Innovationskraft zu sichern.
- **Reduktion von Abhängigkeiten**: Deutschland sollte sich aus der faktischen Dominanz US-amerikanischer Hyperscaler lösen – nicht durch Substitution, sondern durch eine starke eigene Präsenz im Edge-Bereich und klare, offene Schnittstellen zur Cloud.
- **Chancen im industriellen Ökosystem**: Deutschland verfügt über besondere Stärken in Edge Computing, IoT und Embedded-Systemen. Diese Kompetenzen gilt es gezielt in Rechenzentrums- und Cloud-Infrastrukturen zu überführen.
- **Komplexitätsbewältigung**: Heterogene Hardwarearchitekturen erfordern interoperable, hardware-bewusste Software-Stacks, standardisierte Orchestrierungswerkzeuge und hochqualifizierte Fachkräfte. Förderpolitische Impulse sind notwendig, um diese Strukturen nachhaltig zu etablieren.

3 - Welche Rahmenbedingungen sollten aus Ihrer Sicht wie verändert werden, um Rechenzentrumsinvestitionen zu fördern und Innovation zu ermöglichen?

Um Rechenzentrumsinvestitionen zu fördern und Innovation zu ermöglichen, sollten folgende Rahmenbedingungen geschaffen werden:

- **Energiepolitische Flankierung**: Sicherstellung wettbewerbsfähiger Strompreise durch gezielte Entlastungen sowie durch die priorisierte Integration von Rechenzentren als **intelligente Lasten** in Energiesysteme auf Basis erneuerbarer Energien.
- **Planungs- und Investitionssicherheit**: Langfristige regulatorische Klarheit zu Energieabgaben, CO_2-Bepreisung und Netzintegration, um Investitionen kalkulierbar zu machen.
- **Förderung von Nachhaltigkeit**: Unterstützung von Rechenzentren bei innovativen Energiekonzepten (z. B. flexible Lastverschiebung, Abwärmenutzung) als Voraussetzung für klimaneutralen Betrieb.

4 - Welche Rolle sollte der Staat bei der Entwicklung einer souveränen und resilienten Recheninfrastruktur einnehmen?

Der Staat sollte bei der Entwicklung einer souveränen und resilienten Recheninfrastruktur folgende Rollen aktiv übernehmen:

- **Innovationsfördernder Auftraggeber**: Öffentliche Ausschreibungen sind so zu gestalten, dass auch KMU und spezialisierte Anbieter leichteren Zugang erhalten und Wettbewerb sowie Vielfalt gestärkt werden.
- **Energiepolitischer Ermöglicher**: Investitionen in große Stromspeicherinfrastrukturen sollten flankiert werden, damit Rechenzentren günstige und verlässlich verfügbare erneuerbare Energie nutzen können.
- **Nachhaltigkeitsmotor**: Durch Förderung modularer, wiederverwendbarer Systeme und die Reduktion von „embodied CO_2“ in Hardware und Infrastruktur kann der Staat ökologische Leitplanken setzen und gleichzeitig Innovationsanreize geben.

5 - Gibt es konkrete Maßnahmen oder Best Practices aus Ihrer Praxis/Erfahrung, die in die Strategie aufgenommen werden sollten?

Folgende konkrete Maßnahmen und Best Practices sollten in die nationale Rechenzentrumsstrategie aufgenommen werden:

- **Einsatz modularer, heterogener Hardware**: Durch Wiederverwendbarkeit von Komponenten (modulare Serverarchitektur) und geringeres „embodied CO_2“ lassen sich Umweltwirkungen reduzieren und Investitionen effizienter nutzen.
- **Effizienzgewinne durch Workload-Spezialisierung**: Heterogene Architekturen ermöglichen 2–10x höhere Effizienz bei zielgerichteten Workloads, führen zu besserer Ressourcenauslastung und senken die Total Cost of Ownership (TCO).
- **Souveränität und Nachhaltigkeit**: Der gezielte Einsatz spezialisierter Hardware steigert Rechenleistung pro Watt und pro Euro, stärkt die technologische Unabhängigkeit und senkt Emissionen.
- **Komplexitätsmanagement**: Damit heterogene Infrastrukturen ihre Wirkung entfalten können, sind Investitionen in interoperable Software-Stacks, standardisierte Orchestrierungswerkzeuge und Qualifizierung von Fachkräften zwingend erforderlich.
- **Förderung spezialisierter Edge-Beschleunigerprojekte**: Europäische Förderinstrumente sollten gezielt Initiativen unterstützen, die kompakte, energieeffiziente Beschleuniger für Edge-Szenarien entwickeln und in heterogene Architekturen integrieren. Solche Projekte ermöglichen dynamisches Workload-Splitting und energieproportionalen Betrieb, sind zugleich leichter zugänglich für KMU und Forschungseinrichtungen und stärken so die Breite der Innovationsbasis.
- **Förderung von Toolchains für Workload-Aufteilung**: Neben der Hardwareentwicklung sollten Software-Toolchains gefördert werden, die eine automatische Partitionierung, Orchestrierung und Laufzeitoptimierung von Workloads auf verschiedene Beschleunigerklassen erlauben. Sie sind entscheidend, um Effizienzpotenziale praktisch nutzbar zu machen und die Komplexität heterogener Systeme für Betreiber zu reduzieren.

6 - Referenzen

[1] European Commission. (2025, 9. April). *The AI Continent Action Plan* (COM(2025)165). Brüssel: Europäische Kommission.
URL: https://digital-strategy.ec.europa.eu/en/library/ai-continent-action-plan

[2] Hintemann, R., Hinterholzer, S., & Progni, K. (2024). *Bitkom-Studie Rechenzentren in Deutschland: Aktuelle Marktentwicklungen – Stand 2024*. Berlin: Bitkom / Borderstep Institut für Innovation und Nachhaltigkeit.
URL: https://www.borderstep.de/publikation/hintemann-r-hinterholzer-s-progni-k-2024-bitkom-studie-rechenzentren-in-deutschland-aktuelle-marktentwicklungen-stand-2024-berlin-bitkom/

[3] Hintemann, R., Hinterholzer, S., & Konrat, F. (2024, 18.–20. Juni). *Server Stock Data — A Basis for Determining the Energy and Resource Requirements of Data Centres*. In *Proceedings of Electronics Goes Green 2024+ (EGG)*. EGG, Berlin.
DOI: 10.23919/EGG62010.2024.10631194
URL: https://www.borderstep.de/wp-content/uploads/2024/06/Server-stock-data-EGG2024.pdf

[4] Power Electronics News. (2024). *Schneider Electric predicts substantial energy consumption for AI workloads globally*.
URL: https://www.powerelectronicsnews.com/schneider-electric-predicts-substantial-energy-consumption-for-ai-workloads-globally/

[5] M2DC – Modular Microserver DataCentre. URL: https://m2dc.eu/en/

[6] Oleksiak, A., Kierzynka, M., Piatek, W., vor dem Berge, M., Christmann, W., Krupop, S., Porrmann, M., Hagemeyer, J., Griessl, R., Peykanu, M., Tigges, L., Rosinger, S., Schlitt, D., Pieper, C., Janssen, U., Rauchfuss, H., Agosta, G., Barenghi, A., Brandolese, C., Fornaciari, W., Pelosi, G., … Ceva, L. (2018). *M2DC — A Novel Heterogeneous Hyperscale Microserver Platform*. In: *Hardware Accelerators in Data Centers*. Springer, Cham. DOI: 10.1007/978-3-319-92792-3_6. URL: https://www.m2dc.eu/media/filer_public/51/f4/51f41c26-f297-4a9f-9973-d9ffb15f110c/2017_m2dc__modular_microserver_datacentre_with_heterogeneous_hardware.pdf

[7] VEDLIoT project. URL: https://vedliot.eu/

[8] Mika, K., Griessl, R., Kucza, N., Porrmann, F., Kaiser, M., Tigges, L., Hagemeyer, J., Trancoso, P., Azhar, M. W., Qararyah, F., Zouzoula, S., Ménétrey, J., Pasin, M., Felber, P., Marcus, C., Brunnegard, O., Eriksson, O., Salomonsson, H., Ödman, D., Ask, A., Casimiro, A., Bessani, A., Carvalho, T., Gugala, K., Zierhöffer, P., Latosinski, G., Tassemeier, M., Porrmann, M., Heyn, H.-M., Knauss, E., Mao, Y., Meierhöfer, F. (2023). *VEDLIoT — Next generation accelerated AIoT systems and applications*. 20th ACM International Conference on Computing Frontiers (CF '23), Bologna, Italy. DOI: 10.1145/3587135.3592175, URL: https://arxiv.org/abs/2305.05388

[9] CAPE project – European Open Compute Architecture for Powerful Edge.
URL: https://cape-project.eu/

[10] *CAPE – European Open Compute Architecture for Powerful Edge*. CAPE Projekt-Arbeitsdokument „Architectural Blueprint for Powerful Edge". September 2025.
URL: https://cape-project.eu/wp-content/uploads/2025/09/DSD_2025_CAPE-European_Open_Compute_Architecture_for_Powerful_Edge.pdf

[11] Christmann Informationstechnik + Medien GmbH & Co. KG. *CoolEdgeHPC – Modulare, heterogene Edge-to-HPC-Microserverplattform mit leiterplatten-integrierter Entwärmung*. Projektbeschreibung. Ilsede: Christmann. URL: https://christmann.info/cooledgehpc/ , https://www.elektronikforschung.de/projekte/cooledgehpc

[12] https://www.uni-bielefeld.de/kognitronik-sensorik/

[13] https://embedded.christmann.info/products/