Die folgenden Antworten beziehen sich hauptsächlich auf Hochleistungsrechner/Supercomputer für numerische Simulationen und KI-Anwendungen, wie z.B. den kürzlich eingeweihten „JUPITER“.

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

- Niedrigschwelliger Zugang zu Systemen für die Industrie/Forschung mit minimaler Bürokratie (zumindest, wenn nur begrenze Rechenkapazitäten im Bereich von vier- bis fünfstelligen Knotenstunden benötigt werden), Antragsverfahren dürfen nicht (mehr) >0.5 Jahre brauchen

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

- Herausforderungen: Stromkosten machen großen/größten Teil der Kosten eines Rechenzentrums im Laufe seines Lebenszyklus aus, Stromkosten in Deutschland im Vergleich zu Rechenzentren in anderen Ländern weder ökonomisch noch ökologisch wettbewerbsfähig, auch aus geographischen Gründen (Beispiel: Supercomputer „LUMI“ in Kajaani, Finnland wird zu 100 % mit Wasserkraft betrieben), für Nutzer hat es hingegen kaum bis keine Vorteile, wenn der Supercomputer in Deutschland steht
- Chancen: großes Potential an sinnvollen Nutzern, die bisher noch nicht (ausreichend) Supercomputer/Rechenzentren in ihr Geschäftsmodell integriert haben (insb. Mittelstand)

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

- Finanzielle Förderung im Rahmen von Forschungsinitiativen (wie es sie aktuell schon gibt, z.B. „EuroHPC“)

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

- Häufig gibt es eine große Diskrepanz zwischen den maximalen Leistungsdaten, die von Hardwareherstellern (Intel/AMD/NVIDIA) angegeben werden und die sich politisch zwar gut vermarkten lassen („Deutschlands neuer Supercomputer kann X Berechnungen pro Sekunde durchführen“), allerdings in realen Anwendungen niemals erreicht werden. Auch das Errichten von großen Supercomputern mit tausenden von GPU-Knoten (wie z.B. „JUPITER“) macht häufig den Eindruck, dass es dabei auch um Prestige geht. Es sollte aber nicht der Blick für die tatsächlich gewünschten Anwender verloren gehen die die Systeme volkswirtschaftlich produktiv nutzen, wie z.B. die Industrie, die in ihren Simulationen häufig weiterhin auf CPUs setzen und nur einen Bruchteil des Supercomputers nutzen (häufig sogar nur einen Rechenknoten) und daher ein Cluster mit tausenden Knoten für diese Nutzer keinen Vorteil bringt im Vergleich zu zwei halb so großen Clustern. Zwar bringen große Cluster sicherlich auch finanzielle Skalierungsvorteile, aber auch Clusterrisiken (Ausfall/Angriff auf das Rechenzentrum betrifft gleich eine große Menge an Anwendern)
- Was den vorangegangenen Punkt noch verschärft: die meisten Nutzer erreichen aufgrund mangelnder Erfahrung und daraus resultierender fehlerhafter Bedienung mit

ihren Anwendungen nur einen Bruchteil der für sie maximal erreichbaren Leistung, daher sollten entsprechende Schulungen immer mitgedacht werden

Für Rückfragen: samuel.kemmler@bam.de