



Technologie Trends

Server, Speicher, Netzwerk

www.bitkom.org

bitkom

Danksagung

Zum Gelingen dieses Leitfadens hat der gesamte Bitkom Arbeitskreis »Server, Storage, Networks« beigetragen. Insbesondere danken wir folgenden Mitgliedern (in alphabetischer Reihenfolge):

- Arne Lehfeldt (EMC Deutschland GmbH)
- Frank Beckereit (Dimension Data Germany AG & Co. KG)
- Georg Mey (NetApp Deutschland GmbH)
- Georgios Rimikis (Hitachi Data Systems GmbH)
- Ingolf Wittmann (IBM Deutschland GmbH)
- Peter Dümig (Dell Halle GmbH)
- Thomas Arenz (Samsung Semiconductor Europe GmbH)
- Ulrich Hamm (Cisco Systems GmbH)
- Claus Emrich (Microsoft)

1 Über diese Publikation

Bitkom stellt regelmäßig einen bewusst sehr kurz gehaltenen Überblick über die aktuellen Trends bei Server-, Speicher- und Netzwerktechnologien im professionellen Bereich zusammen. Wir wollen mit der Trend-Übersicht die einzelnen Themen nicht inhaltlich erklären, denn dazu gibt es bereits zahlreiche andere Quellen. Unser Ziel ist es, die wesentlichen Trends aufzuzeigen und Entscheidern in Unternehmen helfen, die aktuellen technischen Entwicklungen in Bezug auf ihre jeweilige geschäftliche Relevanz besser einschätzen zu können.

Die Trends werden etwa zweimal pro Jahr überprüft und gegebenenfalls überarbeitet.

2 Servertechnologie und Computing

2.1 Container

Container, bekannt vor allem durch Projekte wie »Docker«, sind vereinfacht gesagt eine Möglichkeit, Applikationen unabhängig vom Betriebssystem und auf Basis von gemeinsamen Templates (sog. »Images«) zu erstellen. Ein Container besteht dabei im Wesentlichen aus drei Schichten: einem »Image« (dem eigentlichen Template, ähnlich einer Klasse bei Objektorientierter Programmierung), einem Container (einer Art funktionspezifischen read-only-Kopie des Templates) sowie einer anwendungsspezifischen (bzw. containerspezifischen) Datenschicht in der die Daten mit Schreibrecht abgelegt werden. Das gesamte Konzept ist relativ komplex und wird hier nur sehr unzureichend beschrieben, ermöglicht aber sehr elegante DevOps-Konzepte und Anwendungsportabilität.

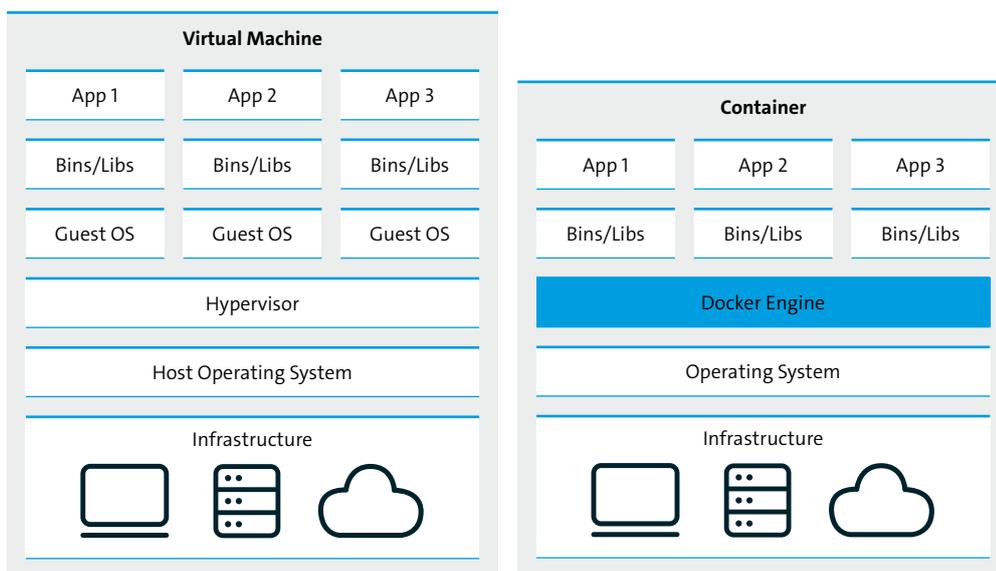


Abbildung 1: Virtual Machines vs. Docker

2.2 Microservices

Anwendungen bestehen aus vielen einzelnen Funktionsblöcken, die untereinander kommunizieren. Werden solche Funktionen ausreichend abgegrenzt definiert und als eigene Anwendung erstellt, lassen sich diese Funktionsblöcke auch für unterschiedliche Anwendungen gemeinsam nutzen (im weitesten Sinne ähnlich der Vererbung bei der Objektorientierten Programmierung). Ein sehr, sehr stark vereinfachtes Vergleichsmodell wäre ein Service welcher Statusmails verschickt, der sowohl von einem Backupprogramm als auch von der Fahrstuhlsteuerung oder der Zisternensteuerung verwendet werden kann, ohne in jeder Anwendung neu geschrieben werden zu müssen. Microservices können dabei in beinahe beliebiger Weise (z. B. als Unix Daemon, Windows Service, Webservice, etc.) realisiert werden.

2.3 Edge Computing

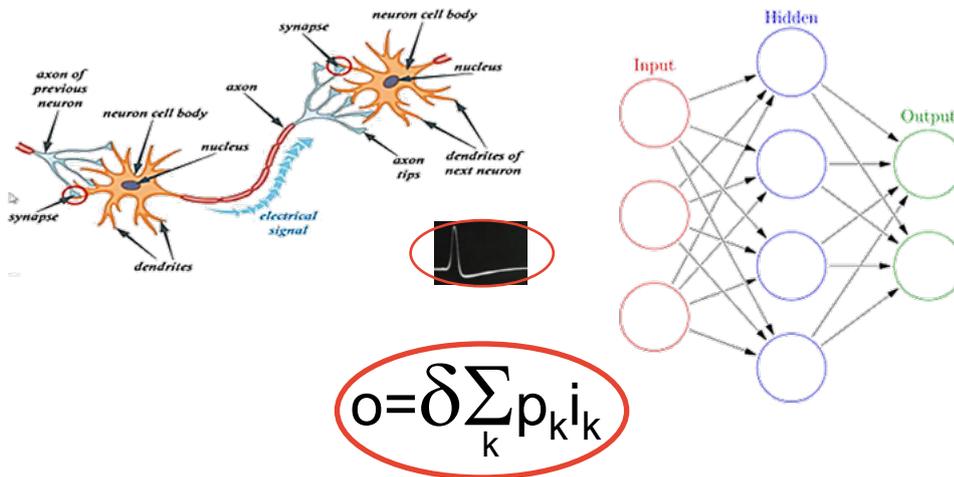
Milliarden von IoT Geräten erzeugen enorm große und meistens unstrukturierte Datenmengen, die schnell verarbeitet werden müssen, um daraus wertvolle Erkenntnisse zu gewinnen. Der Ansatz von »Edge Computing« besteht darin, den größten Teil der zu erfüllenden Aufgaben nicht in einem zentralen Rechenzentrum zu bewältigen, sondern an dem Ort zu erledigen, an dem die Daten entstehen, sofern der Anwendungsfall dies erlaubt und die informationstechnischen Anforderungen erfüllt werden können.

2.4 Neuromorphic Computing

Neuromorphic Computing beschäftigt sich mit biologisch inspirierter Physik, Mathematik, Informatik und Ingenieurwissenschaften zur Entwicklung neuronaler Hardwarearchitekturen, welche sich von Turing Modellen grundsätzlich unterscheiden. Diese neuen Technologien stellen modulare Erweiterungen zu einem klassischen von-Neumann-Rechner dar und sollen effizienter neuronale Netzwerkentwicklung unterstützen bzgl. Geschwindigkeit, Energieeffizienz und Programmierbarkeit.

Neuromorphic Processing Units (NPU) sind dem menschlichen Gehirn in der Funktionsweise nachempfunden und können als einzelne Einheiten in mobilen Geräten z.B. für Sprach- oder Bilderkennung oder in großen Clusterumgebungen für Mustererkennung, Big Data und KI eingesetzt werden. Der wesentliche Unterschied zu klassischen Prozessoren ist die hohe Vernetzung innerhalb der NPUs mit Möglichkeiten der Rückkoppelung.

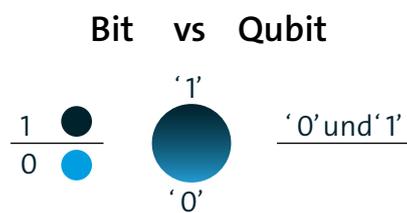
Eine Erweiterung von Supercomputern durch Neuromorphic Module ermöglicht es, Systemarchitekturen zu entwickeln, um exascale-Leistung zu erreichen, wo ein klassischer Exaflop Rechner alleine am Stromverbrauch scheitern würde.



Das Bild¹ stellt die Analogie der Signalverarbeitung in biologischen Nervenzellen und in neuronalen Netzen dar. Ein über eine Synapse eingespeistes Eingangssignal wird über eine oder mehrere Verarbeitungsschichten in ein Ausgangssignal transformiert. Die Transformation kann man sich vereinfacht als eine mit individuellen Faktoren gewichtete Überlagerung aller Eingangsgrößen vorstellen. Die kontrollierte Anpassung der Gewichte bildet den Lernprozess des neuronalen Netzes ab.

2.5 Quantum Computing

In der IT wird immer mehr das Thema Quantum Computing diskutiert, nachdem die ersten Systeme für die Privatwirtschaft in greifbare Nähe geraten sind.



Jedes Qubit kann in einem Zustand von sowohl »0« als auch »1« sein, so dass N Qubits 2^N Zustände darstellen können.²

Ein Quantencomputer oder Quantenrechner basiert auf den Gesetzen der Quantenmechanik und ist mit einer klassischen Von-Neumann-Architektur nicht vergleichbar. Ähnlich wie in einem klassischen Rechner, werden Quantenbits (Qubits) als kleinste Informationseinheit verwendet. Aufgrund des sogenannten Superpositionsprinzips kann ein Qubit während der Rechenzeit eine

1 Quelle: IBM Research Almaden

2 Quelle: IBM Research Yorktown

beliebige Überlagerung der Zustände 0 und 1 annehmen. Durch die Verschränkung von N Qubits können 2^N Zustände parallel dargestellt und verarbeitet werden. Damit können Problembereiche adressiert werden, welche ein klassischer Computer in einer überschaubaren Zeit nicht mehr handhaben kann. Man geht davon aus, dass ein universeller Quantencomputer mit 50 Qubits nicht mehr von einem heutigen klassischen Rechner simuliert werden kann.

Ein Quantencomputer wird aber nicht die klassischen von-Neumann-Rechner ablösen. Es gibt Problembereiche wie z. B. ERP, Webhosting, Visualisierung, in denen ein Quantencomputer wohl keine Vorteile bringt. Wenn es aber um die Simulation von Molekülen, Optimierungsproblemen wie den »Traveling Salesman«, Machine Learning oder aber auch um die Suche in großen unstrukturierten Datenmengen geht, kann Quantum Computing seine exponentiellen Skalierungsfähigkeiten ausspielen.

Quantencomputer werden immer eng mit klassischen Von-Neumann-Rechnern integriert sein und sich gegenseitig ergänzen.

Quantum Computing wird zusammen mit anderen Quantentechnologien, wie der Quantenkryptographie, zu einer »Revolution« in der Informationsverarbeitung führen und gänzlich neue Themengebiete öffnen.

2.6 Serverless Computing

Serverless Computing beschreibt eine Plattform zur Entwicklung, Bereitstellung, Ausführung und Skalierung von Anwendungen. Bisher wurden Anwendungen von Entwicklern in mehr oder weniger großen Blöcken erstellt, und als Anwendungspaket zum Beispiel auf einem Server, in einer virtuellen Maschine oder einem Container zur Ausführung gebracht. Kleinere Anwendungseinheiten werden z.B. via Microservices realisiert. Entwickler oder Verantwortliche für den Anwendungsbetrieb mussten sich dabei um alle Aspekte der nötigen Infrastruktur, wie z.B. die Netzwerkanbindung, die Beschaffung des Servers oder der virtuellen Maschine, die Wartung des Betriebssystems oder auch die Pflege der Container Images kümmern.

Serverless Anwendungen machen diese Abhängigkeiten und Aufwände für die Bereitstellung und den Betrieb einer Infrastruktur obsolet. Um diese radikale Vereinfachung zu ermöglichen, bietet sich für die Entwicklung und Ausführung von Serverless Anwendungen eine Cloud Plattform an. Anwendung werden dabei in kleinste Funktionsblöcke zerlegt, die meist Event-basiert agieren.

Dieser Anwendungstyp ermöglicht es Entwicklern, den Quellcode bzw. die Funktionseinheiten ihrer Anwendung direkt auf die Serverless Computing Schicht des Cloud Anbieters aufzubringen und auszuführen. Es steht also eine vollumfängliche, automatisch skalierende, hochverfügbare Laufzeitumgebung zur Verfügung welche nicht vom Entwickler oder dem IT-Betrieb verwaltet werden muss. Gängig ist außerdem für die Ausführung von Anwendungen auf einer Serverless Plattform eine nutzungsorientierte Abrechnung.

Serverless ist aus Nutzersicht als Mittelweg aus Platform-as-a-Service (PaaS) und Software-as-a-Service (SaaS) zu sehen. Es erlaubt, im Gegensatz zur Nutzung einer fertigen Softwarelösung (SaaS), die Entwicklung und Ausführung eines komplett selbsterstellten Quellcodes.

Natürlich benötigt Serverless Computing aus Sicht des (Cloud)-Plattform Anbieters weiterhin Server, Speichersysteme und Netzwerk. Der Begriff »Serverless Computing« ist daher aus Sicht des Nutzers, im speziellen des Software Entwicklers, zu sehen. Aus seiner Sicht ist die gesamte zugrunde liegende Infrastruktur verborgen und außerhalb seiner Kontrolle.

3 Speichertechnologie

3.1 SSD

Ein Flash-Speicher (oder SSD bzw. Solid State Disk) ist ein nichtflüchtiges Speichermedium auf der Basis von NAND (heute) und zukünftig auch anderen Technologien (wie u.a. Intel 3D XPoint).

In den nächsten Jahren wird Flash-Speicher die Festplatte als Standard-Medium im Tier 1 ablösen, vermutlich bis 2020 die Festplatte beim Verkauf überholen und spätestens ab 2025 auch in allen anderen Bereichen fast komplett verdrängt haben.

Damit gehen einher massive Geschwindigkeitszuwächse, niedrigere Latenz, deutlich höhere Kapazitäten, niedrigerer Energieverbrauch und auch niedrigere Ausfallraten. Diese Verschiebung ist bereits im Gange und wird nur durch die momentan noch höheren Kosten bei Flash-Medien gedämpft. Analysen zufolge ist heute bereits eine Gleichheit der Kosten eines Flash-Speichers mit 15k SAS HDDs erreicht und im Jahr 2018 mit 10k SAS HDDs. Daher ist danach ein nochmals stärkerer Anstieg der Nutzung zu erwarten. Betrachtet man die TCO über 4 Jahre sind SSDs heute bereits günstiger als HDDs und der Preis fällt schneller als bei Festplatten. Die noch für dieses Jahr angekündigten SSDs mit Kapazitäten bis 60 TB werden diesen Trend weiter beschleunigen und den Vorteil bei Kapazität, Platzbedarf und Stromverbrauch weiter ausbauen.

3.2 NVMe

NVM Express (kurz NVMe) ist ein erstmals im Jahr 2011 veröffentlichtes Protokoll, um nichtflüchtige Massenspeicher (engl. non-volatile memory, kurz NVM) über z. B. PCI Express mit dem Prozessor zu verbinden, ohne dass dafür herstellerspezifische Treiber nötig wären. NVM Express wurde von Grund auf für Speicher mit geringer Latenz und parallelen Zugriffen entwickelt und löst damit etablierte Technologien wie SCSI und SAS/SATA ab.

Aktuell wird am Standard für NVMe über Fabrics gearbeitet. Damit können auch Fibre Channel, Infiniband, Omni-Path, CAPI oder Ethernet statt PCI Express genutzt werden, um so die mögliche Entfernung zwischen Server und Storage zu erhöhen.

3.3 Persistent Memory

Persistent Memory ist eine neue Technologie, welche die Lücke zwischen SSDs und dem Arbeitsspeicher schließt. Die sogenannten NVDIMM Module sind kompatibel zu normalem DDR4 Arbeitsspeicher und können alternativ in einen Teil der vorhandenen Steckplätze eingebaut werden (daher »Memory«). Der Inhalt bleibt auch bei dem Herunterfahren des Systems erhalten (daher »Persistent«). Die leistungsfähigere Anbindung an den Prozessor erlaubt eine deutliche Verringerung der Latenz und erhöht somit die Verarbeitungsgeschwindigkeit datenintensiver Arbeitslasten .

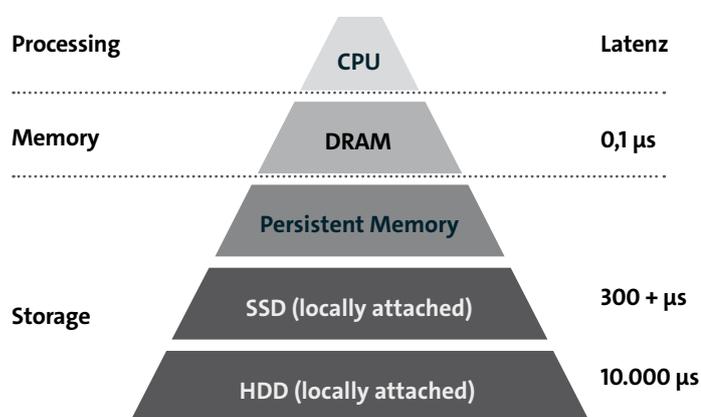


Abbildung 2: Speicherhierarchie unter Einbezug von Persistent Memory

Ideale Anwendungsbereiche sind Datenbanken, Big Data, Software Defined Storage, RAID Cache, In-Memory Datenbanken oder Speicher für Metadaten in Servern.

4 Netzwerktrends

4.1 Overlay-Netze

In den letzten Jahren haben sogenannte Overlay-Netze im Markt Einzug gehalten. Diese Overlay-Netze werden auch als Software Defined Networking (SDN) bezeichnet und stellen eine virtuelle Ebene dar, um über die physikalischen Grenzen von Geräten hinweg zu skalieren und die Bedienbarkeit zu vereinfachen.

Die Overlay-Netze wurden bei den public Cloud Anbietern entwickelt und werden dort bereits eingesetzt. Nun verbreitet sich diese Technik auch immer stärker in on-premises Umgebungen im Markt.

Auf der Basis von SDN lassen sich Skripte und Applikationen zur Automatisierung von Netzwerken wesentlich leichter anwenden. Netzwerke werden nicht mehr per Command Line Interface (CLI) konfiguriert, sondern per Application Programming Interface (API) programmiert. APIs stellen neue Schnittstellen bereit, die die Bedienbarkeit um ein Vielfaches erleichtern. Durch die Einführung von SDN gibt es im Normalfall eine Clusteradresse oder einen zentralen Controller, der die Hoheit über das gesamte Netzwerk besitzt und auch zentral angesprochen werden kann.

Dadurch ist auch das API zentralisiert, und Skripte sowie Anwendungen lassen sich zentralisiert im gesamten Netzwerk ausgeführt. Aufgrund dieser Eigenschaft bekommt man im Netzwerk sehr leicht einheitliche replizierbare Konfigurationen, bei denen manuelle Fehler nahezu ausgeschlossen werden können.

Die Programmierbarkeit bringt einen weiteren erheblichen Vorteil mit sich. Jedes Kommando und jede Funktionalität im Netzwerk ist per API ausführbar. Je nach Benutzerkreis lassen sich somit auch sogenannte Self-Service Portale etablieren, die dem Anwender die Möglichkeit bieten, in einem eingegrenzten Umfang selbst Netzwerke zu provisionieren.

Aber es gibt auch einen Nachteil von SDN Lösungen. Es wird immer schwieriger, im Fehlerfall zu agieren. Macht das »Overlay« das richtige? Gibt es eventuell im »Underlay« (also auf der Physik) Probleme, die das Overlay nicht erkennt? Funktionieren die Regelwerke genauso, wie sie gedacht sind?

Diese Probleme entstehen vor allem durch die erhöhte Komplexität der Netzwerke und können durch ergänzende Lösungen, die im Markt immer mehr angenommen werden, behoben werden.

4.2 Network Analytics

In modernen Netzwerken gibt es zahlreiche Kommunikationsbeziehungen zwischen Applikationen. Network Analytics befasst sich mit der Analyse von Netzwerkverkehrsdaten und Statistiken, um Trends und Muster zu erkennen. Das Erkennen der Kommunikationsbeziehungen kann aus mehreren Gründen vorteilhaft sein. So können beispielsweise Anomalien erkannt werden, die auf das Eindringen eines Hackers hinweisen. Ebenso können mögliche Performance-Probleme an einzelnen Netzelementen erkannt werden, und als Abhilfe könnte der Netzwerkverkehr ggf. umgeleitet werden.

Um Anomalien, also Änderungen im typischen Kommunikationsverhalten von Anwendungen, im Netzwerk zu erkennen und Gegenmaßnahmen zu ermöglichen, wird über einen gewissen Zeitraum, meist einige Wochen oder auch Monate, beobachtet, welche Ports und Protokolle typischerweise von den Anwendungen im Netzwerk genutzt werden. Auf dieser Grundlage wird ein Regelwerk automatisiert erstellt, welches fortan als Referenz dient. Sollten in der Folge abweichende Kommunikationsbeziehungen auftreten, können diese vom Netzwerk blockiert oder an den Administrator gemeldet werden.

Für die Analyse der Netzwerkdaten müssen Telemetriedaten aus den Netzwerkkomponenten gesammelt werden. Dies kann zum einen über Software erfolgen, die in den Netzelementen installiert wird, z.B. in Servern. Zum anderen können einige Netzwerkelemente, z.B. Router und Switches, bereits über ihre Hardware die erforderlichen Telemetriedaten zur Verfügung stellen. Diese Methode hat den Vorteil gegenüber bestehenden Techniken, dass sie sehr performant ist und damit alle Flows erfasst werden können, was bei Softwarelösungen nicht immer gelingt.

Bitkom vertritt mehr als 2.600 Unternehmen der digitalen Wirtschaft, davon gut 1.800 Direktmitglieder. Sie erzielen allein mit IT- und Telekommunikationsleistungen jährlich Umsätze von 190 Milliarden Euro, darunter Exporte in Höhe von 50 Milliarden Euro. Die Bitkom-Mitglieder beschäftigen in Deutschland mehr als 2 Millionen Mitarbeiterinnen und Mitarbeiter. Zu den Mitgliedern zählen mehr als 1.000 Mittelständler, über 500 Startups und nahezu alle Global Player. Sie bieten Software, IT-Services, Telekommunikations- oder Internetdienste an, stellen Geräte und Bauteile her, sind im Bereich der digitalen Medien tätig oder in anderer Weise Teil der digitalen Wirtschaft. 80 Prozent der Unternehmen haben ihren Hauptsitz in Deutschland, jeweils 8 Prozent kommen aus Europa und den USA, 4 Prozent aus anderen Regionen. Bitkom fördert und treibt die digitale Transformation der deutschen Wirtschaft und setzt sich für eine breite gesellschaftliche Teilhabe an den digitalen Entwicklungen ein. Ziel ist es, Deutschland zu einem weltweit führenden Digitalstandort zu machen.

**Bundesverband Informationswirtschaft,
Telekommunikation und neue Medien e.V.**
Albrechtstraße 10 | 10117 Berlin

Christian Herzog | Bereichsleiter IT-Infrastruktur & Kommunikationstechnologien
T 030 27576-270 | c.herzog@bitkom.org

www.bitkom.org

bitkom