

# Die Top-Clouds von heute setzen auf Intel

## Große Anbieter von Cloud-Diensten (CSPs, Cloud Service Providers) nutzen für die Hyperskalierung Intel® Architektur.

Manager von Rechenzentren kennen die Herausforderungen: Skalierung der Infrastruktur zur Bewältigung wachsender Datenmengen, Optimierung von Systemen für anspruchsvolle neue Workloads wie künstliche Intelligenz (KI), Schutz von Anwendung und Daten vor immer neuen Bedrohungen und das ständige Bemühen, die Performance auf allen Ebenen zu verbessern. CSPs stehen vor denselben Herausforderungen was Hyperskalierung betrifft. Ihre Lösungen entwickeln sie häufig in Zusammenarbeit mit Intel.

Intel ist seit Jahrzehnten Vorreiter bei der Hyperskalierung von Cloud-Diensten. Dank gemeinsamer Entwicklungsarbeit und geschäftlicher Beziehungen mit Top-CSPs auf der ganzen Welt konnte Intel Generation maßgeschneiderter Hardware liefern, die für die Cloud-Skalierung entworfen und optimiert wurde. Viele der ursprünglich für Hyperscaler entwickelten Funktionen wurden seither in das datenzentrierte Portfolio der Intel® Techniklösungen aufgenommen und werden nun in Rechenzentren aller Größen genutzt.

Dieser Artikel beschreibt einige der wichtigsten Intel® Techniklösungen, die bei den größten CSPs der Welt im Einsatz sind. Diese Informationen sollten auf zweierlei Weise nützlich sein:

- Wenn Workloads in eine Public Cloud migriert werden, helfen diese Informationen, die Vorteile optimierter Intel® Technik zu erkennen und verstehen, die die zu migrierenden Workloads unterstützen, wodurch die CSP-Angebote ermittelt werden können, die den Anforderungen am besten gerecht werden.
- Wenn weiterhin kritische Workloads vor Ort im eigenen Rechenzentrum laufen, dann sollte in Betracht gezogen werden, einige derselben Technologien einzusetzen, die CSPs nutzen, um ihre Probleme bei der Hyperskalierung zu lösen. Sich für Techniklösungen zu entscheiden, die sich in den größten Rechenzentren der Erde bewährt haben, ermöglicht eine solche Infrastrukturkompatibilität. Durch diese können Workloads einfacher je nach Bedarf zwischen eigenem Rechenzentrum und Public Clouds verschoben werden.

Manche Lösungen wie die Intel® Turbo-Boost-Technik und die Intel® Hyper-Threading-Technik (Intel® HT-Technik) lassen sich in beinahe

alle auf Intel basierende Instanzen bei einem CSP integrieren. Andere sind jedoch möglicherweise nur für einen bestimmten Teil dieser Instanzen verfügbar, der identifiziert werden muss. Intel engagiert sich für flächendeckenden Zugang zu Cloud-Diensten, damit Unternehmen herausfinden können, welches Geschäftsmodell für sie am besten geeignet ist.

Die in diesem Artikel besprochene Intel® Technik ist in folgende Kategorien unterteilt:

- **Intelligente Netzwerkkonnektivität.** CSPs sind auf leistungsstarke Netzwerkkonnektivität angewiesen, die flexibel ist und sich an wechselnde Anforderungen und Workloads anpasst, um Daten schneller zu übertragen und eine hohe Rechenleistung aufrechtzuerhalten.
- **Anwendungs- und Datensicherheit.** Unternehmen müssen sich darauf verlassen können, dass ihre Daten und Workloads, die in die Cloud verlagert werden, weiterhin vertraulich und vor Bedrohungen geschützt sind – und sogar davor, dass der Cloud-Anbieter selbst auf die Daten zugreifen kann.
- **KI und High-Performance-Computing (HPC).** CSPs sind zunehmend in der Lage, diese äußerst rechenintensiven Workloads in spezialisierten Instanzen unterzubringen, die auf Intel® Technik basieren.
- **Performancesteigerung.** Von Unternehmen wird dauernd erwartet, mit weniger mehr zu erreichen, indem die Workload-Performance und -Dichte durch Innovationen bei Prozessoren, Arbeits- und Datenspeicher verbessert wird.

In diesen Zeiten des rasanten Wandels benötigen sie daher agile, zuverlässige und skalierbare Lösungen, die Geschäftskontinuität und langfristigen Erfolg sicherstellen. Geschäftliche Anforderungen entwickeln sich schnell weiter und die IT-Infrastruktur muss in der Lage sein, ebenso schnell darauf zu reagieren. Das kann bedeuten, schnell vor Ort zu skalieren oder die Workload-Performance in der Cloud zu optimieren, um wechselnden Anforderungen gerecht zu werden. Intel® Techniklösungen bilden die Basis für Unternehmensrechenzentren





und die Serviceangebote von Public CSPs. Sie bieten Performance, die darauf abgestimmt ist, die Anforderungen der Workloads von heute und morgen zu erfüllen, sowie eine verbesserte Betriebseffizienz, um vom Rechenzentrum zur Cloud und zum Edge zu skalieren. Darüber hinaus sind sie eine zuverlässige technologische Grundlage für problemlose Umstellungen und Migrationen.

Was sind die wichtigsten Intel® Techniklösungen für die Cloud und wie funktionieren sie? Wie werden sie von CSPs genutzt und welchen wirtschaftlichen Wert bieten sie sowohl den CSPs als auch ihren Kunden? Dieser Artikel liefert einen kurzen Überblick über die verschiedenen Techniklösungen, um solche Fragen zu beantworten. Zu verstehen, wie Hyperscaler Intel® Technik einsetzt, kann für Infrastrukturarchitekten und andere Entscheidungsträger in digitalen Unternehmen bei der Entwicklung einer Cloud-Strategie von Nutzen sein. Diese Entscheidungsträger benötigen eine klare Vorstellung davon, wie sich ihre Entscheidungen in Bezug auf Private- und Public-Cloud-Dienste und deren zugrundeliegende Intel® Technik auswirken.

## Leistungsstarke, intelligente Netzwerkkonnektivität

Hyperscaler bevorzugen eine intelligente, skalierbare softwaredefinierte Infrastruktur gegenüber Hardware und Appliances, die spezifisch sind und nur einem einzelnen Zweck dienen. Dafür sind zwei Trends hauptverantwortlich: die Verteilung von Rechenleistung an Remote-Standorte und der Transfer von Rechenzentrumsfunktionen näher ans Edge. Rechenleistung, Datenverarbeitung und Analysen können dadurch beeinträchtigt werden, dass die Netzwerkinfrastruktur Daten nicht schnell genug überträgt, um die zunehmenden Anforderungen an das Netz aufgrund der Traffic-Zunahme zu erfüllen. Software-Defined Networking (SDN) – leistungsstarke Netzwerkkonnektivität, die flexibel ist und sich den wechselnden Anforderungen und Workloads anpasst – ist von entscheidender Bedeutung, um die Rechenleistung optimal zu nutzen. Von Ethernet-Netzwerkkarten und Intel® Silizium-Photonik bis zu Switching-Produkten und -Technik – Intel blickt auf eine lange Geschichte der Innovation bei Netzwerkprodukten zurück. Diese umfassen Hardware, Software und Lösungen, die ein breites Ökosystem ermöglichen, von dem Hyperscaler profitieren.

### Intel® Ethernet Netzwerkcontroller und -adapter

Da sich die Ethernet-Bandbreite von 25 Gbps (Gigabit pro Sekunde) auf 50 bzw. 100 Gbps erhöht hat und zahlreiche

Anwendungen Netzwerkbandbreite beanspruchen, ist der für die Netzwerkschnittstelle erforderliche Verarbeitungs-Overhead zu einem Problem geworden. Verschärft wird dieses Problem noch durch die zunehmende Komplexität moderner Rechenzentrumsnetzwerke, die Techniken zur Performancesteigerung, Virtualisierung und Overlay-Netzwerke unterstützen. Rechenzentrumsbetreiber benötigen zur Erfüllung der dynamischen Anforderungen an das Netzwerk schnelle und intelligente Ethernet-Adapter. Dazu kann auch gehören, einen Teil der Verarbeitung der Infrastruktur-Workloads auf der NIC (Netzwerkkarte; Network Interface Card) zu beschleunigen. CSPs nutzen Intel® Ethernet Adapter aufgrund ihrer bewährten breiten Interoperabilität, wichtigen Leistungsoptimierungen und Agilität.

[Intel® Ethernet Netzwerkcontroller und -adapter](#) unterstützen Geschwindigkeiten von bis zu 100 Gbps. Die Hauptmerkmale von Intel® Ethernet Netzwerkadaptern, von denen Hyperscaler und großen Rechenzentren im Allgemeinen profitieren, sind:

- **Schnellere Paketverarbeitung durch eine verbesserte Performance für die Virtualisierung von Netzwerkfunktionen (NFV, Network Functions Virtualization).** Intel® Ethernet Netzwerkadapter nutzen eine Kombination von Merkmalen zur Hardware- und Software-Beschleunigung. Dazu gehören [Dynamic Device Personalization \(DDP\)](#), das eine anpassbare Paketfilterung ermöglicht, erweiterter Support für das [Data-Plane-Development-Kit \(DPDK\)](#) zur verbesserten Paketweiterleitung und eine hocheffiziente Paketverarbeitung für Cloud- und NFV-Workloads.
- **Verbesserte Traffic-Steuerung zur besseren Erfüllung von Service-Level-Agreements (SLAs).** Die Intel® Ethernet 800 Reihe mit [Application Device Queues \(ADQ\)](#) bietet dedizierte Warteschlangen und regelt den Traffic für die Übertragung von Daten über das Ethernet für kritische Anwendungen. Sie verbessert die Vorhersehbarkeit der Anwendungsreaktionszeit, senkt die Latenz und erhöht den Durchsatz für wichtige Anwendungen wie Datenbanken, den Web-Tier und Caching-Anwendungen. Das hilft Hyperscalern dabei, SLAs besser zu erfüllen.
- **Server-Virtualisierung mit flexibler und skalierbarer I/O-Virtualisierung.** Virtualisierung von Intel sorgt für I/O-Performance und reduziert I/O-Engpässe durch Lösungen wie Virtual Machine Device Queues (VMDq) und Flexible Port Partitioning (FPP), die Single-Root I/O-Virtualisierung (SR-IOV) für den Netzwerk-Traffic mittels virtueller Maschine (VM) nutzen. Dieser Ansatz ermöglicht beinahe native Performance und VM-Skalierbarkeit. Dank der Intel® Virtualisierungstechnik (Intel® VT) bieten Intel® Ethernet Netzwerkadapter herausragende I/O-Performance in virtualisierten Serverumgebungen.

### Smart Network Interface Cards (SmartNICs)

Wie andere Rechenzentrumsbetreiber möchten CSPs Optimierungen für kritische Workloads vornehmen und gleichzeitig effizient bei der Verwaltung von Infrastrukturaufgaben

sein. Wenn sich Infrastrukturaufgaben durch intelligente Netzwerkkomponenten beschleunigen lassen, können Server-CPU's für kritische Workloads entlastet werden.

Hyperscaler nutzen auf Intel® Field Programmable Gate Arrays (FPGAs) basierende [SmartNICs](#) und reduzieren damit den Server-Overhead durch eine Beschleunigung der Aufgaben der Infrastruktur, die üblicherweise von den CPU-Kernen erledigt werden, indem diese nun von den SmartNICs selbst übernommen werden. SmartNICs bieten Performance für Infrastruktur-Workloads. Sie ermöglichen Änderungen der Netzwerktechnologie durch Software-Updates in diesem Bereich. CSPs können SmartNICs dazu nutzen, Host-CPU's zu entlasten, um darauf Endkunden-VMs laufen zu lassen. Ebenso können sie eine SmartNIC dazu nutzen, einen Server als Bare-Metal-Server zu partitionieren, den ein Hosting-Unternehmen einsetzen kann.

SmartNICs, die Intel® FPGAs nutzen, bieten folgende Vorteile für CSPs, die beträchtlich in die Anpassung und Integration von SmartNICs in ihre Infrastrukturen investiert haben:

- Höhere Wirtschaftlichkeit und Skalierbarkeit von Anwendungen, da mehr Host-CPU-Kerne für VMs mit Kunden-Workloads dediziert werden
- Beschleunigte Infrastruktur-Workloads wie zum Beispiel virtualisierte Netzwerkfunktionen (VNF) wie Open vSwitch\*, Sicherheitsprotokolle wie IPsec und Transport Layer Security (TLS) und Datenspeicherprotokolle wie NVMe over Fabrics\* (NVMe-oF\*)
- Die Fähigkeit, Lösungen durch Nutzung der Flexibilität und Programmierbarkeit eines Intel® FPGA zukunftssicher zu machen: Unternehmen können die Funktionalität erweitern oder die Performance der auf den FPGAs laufenden Workloads steigern, damit sie sich veränderten Anforderungen und Normen anpassen
- Niedrige Gesamtbetriebskosten (Total Cost of Ownership, TCO), da dieselbe FPGA SmartNIC je nach Bedarf für verschiedene Workloads bereitgestellt werden kann; zum Beispiel kann die gleiche Hardware zu Spitzenzeiten Sicherheitsanwendungen ausführen und, wenn der Netzwerk-Traffic geringer ist, aktualisiert werden, um darauf Datenanalyse-Workloads laufen zu lassen

## Intel® Tofino™ Switches mit P4-Programmierbarkeit

Hyperscale-Anbieter wollen keine „Black Boxes“ von Herstellern ohne Einblick in den zugrunde liegenden Code und die Funktionalität. Sie benötigen offene Systeme, die sie kontrollieren, debuggen und bei Bedarf umprogrammieren können. Deshalb ist es für sie von Vorteil, wenn sie P4-programmierbare Switches bei ihren softwaredefinierten Netzwerken (SDNs) für Agilität und Skalierbarkeit einsetzen. [Barefoot Networks](#) war ein Vorreiter bei der Entwicklung der Programmiersprache P4 und hat seit der Gründung von P4.org maßgeblich dazu beigetragen, dass das P4-Ökosystem auf weit über 100 Mitglieder angewachsen ist.

Intel unterstützt eine Reihe von Open-Source-Netzwerkbetriebssystemen einschließlich SONiC\*. Das ursprünglich von Microsoft entwickelte SONiC\* ist bei mehreren Hyperscalern im Einsatz.<sup>2</sup> Es basiert auf Linux und läuft auf Switches verschiedener Hersteller und auf anwendungsspezifischen integrierten Schaltungen (ASICs, Application-Specific Integrated Circuits). Intel unterstützt SONiC\*-Code bei den Intel® Tofino™ Switch-ASICs mit 6,4 Tbps (Terabit pro Sekunde) und den Intel® Tofino™ 2 Switch-ASICs mit 12,8 Tbps. Hyperscaler haben dank eines robusten OEM- und ODM-Ökosystems Zugang zu Intel® Tofino™ und Intel® Tofino™ 2 Switches. Zusätzlich zu den White-Box-Switches haben OEMs wie Arista und Cisco programmierbare Switches auf den Markt gebracht, die auf Intel® Tofino™ basieren und SONiC\* unterstützen.

Programmierbare Intel® Tofino™ Switches machen Ethernet-Switches auf der Datentransportebene programmierbar. Dadurch können Nutzer die Funktionalität der Hardware in einem einfachen P4-Programm definieren, auf dem ASIC kompilieren und mit Multi-Tbps-

Geschwindigkeiten laufen lassen. Zum Beispiel ließe sich die Box so einrichten, dass die standardmäßigen Switching- und Routing-Protokolle in größerem Umfang und mit besserer Qualität ausgeführt werden. Sie könnte aber auch durch schnelle Software-Updates um Funktionen wie DDoS-Überwachung (Distributed Denial of Service), Deep Packet Inspection oder Netzwerkadressübersetzung (Network Address Translation, NAT) erweitert werden.

## Unterstützung der Telearbeit während der COVID-19-Pandemie

So wie viele Unternehmen ermöglichte es Intel seinen Mitarbeitern nach dem Ausbruch von COVID-19 schnell, von zu Hause aus zu arbeiten. Diese neu von zu Hause aus arbeitenden Mitarbeiter benötigten sofortigen Zugang zu ihrer Arbeitsumgebung mittels virtueller privater Netzwerke (VPNs). Die Herausforderung bestand darin, auf die notwendige Anzahl von VPNs zu skalieren und dabei die Latenz so niedrig zu halten, dass die Mitarbeiter in der neuen Umgebung produktiv arbeiten können.

Intel IT reagierte darauf mit einer zweigleisigen Strategie, die sowohl firmenintern als auch in einer öffentlichen Cloud umgesetzt wurde. Beide Initiativen setzten auf Intel® Ethernet Netzwerkadapter, die SR-IOV unterstützen – eine Spezifikation, die es einem PCIe\*-Gerät ermöglicht, als mehrere physische PCIe\*-Geräte zu erscheinen –, um eine große Anzahl von VPN-Appliances mit vorhersehbar niedriger Latenz zu skalieren.

Die Public Cloud ließ sich am schnellsten skalieren und trug während der ersten beiden Wochen die Hauptlast. Intel wählte die Instanzen sorgfältig aus, um Intel® Ethernet Netzwerkcontroller mit SR-IOV-Unterstützung und skalierbare Intel® Xeon® Prozessoren einzubinden. Laut den Aufzeichnungen der ersten beiden Wochen waren diese VMs in der Lage, VPN-Nutzer bei etwa 90 Prozent CPU-Auslastung zu unterstützen.

**Intel IT konnte dank der für SR-IOV optimierten Intel® Ethernet Netzwerkadapter die Anzahl der Nutzer pro VPN-Server bei gleichbleibender Latenz um 50 Prozent erhöhen.<sup>1</sup>**

Firmenintern wurden Server mit verschiedenen Intel® Prozessoren eingesetzt, von denen manche schon älter waren. Es wurde dabei Ausrüstung genutzt, die sofort wiederverwendbar war. Einige Systeme, die nicht mit Intel® Ethernet Netzwerkcontrollern ausgestattet waren, ließen sich nicht auf die erforderliche Netzwerk-/VPN-Kapazität skalieren und lieferten inkonsistente Latenzen. Intel IT ersetzte die NICs in diesen Systemen durch Intel® Ethernet Netzwerkadapter, die SR-IOV unterstützen. Diese ließen sich so skalieren, dass die Nachfrage nach einer großen Anzahl von Nutzern mit gleichbleibend niedriger Latenz erfüllt werden konnte.

Da sowohl die Public-Cloud-Instanzen als auch die firmeninternen Server SR-IOV nutzten, wurde der Lastausgleich zwischen den beiden vereinfacht. Ein größerer Teil der Arbeitslast wurde zurück in die firmeneigene Cloud verlagert, als diese Systeme online gingen.

## Intel® Silizium-Photonik

[Intel® Silizium-Photonik](#) bietet schnelle Glasfaserverbindungen über lange Distanzen für riesige lagerhallengroße Hyperscale-Rechenzentren, bei denen die schnelle Übertragung von Daten auf Server über größere Distanzen entscheidend ist, damit sie schnell verarbeitet oder analysiert werden können. Innerhalb von Rechenzentren verbinden optische Verbindungen Switches

durch ein komplexes Netzwerk von Glasfaserkabeln und optischen Transceivern. Server und Datenspeicher müssen sich nicht mehr physisch nahe sein und können unabhängig voneinander skaliert werden, basierend auf verschiedenen Anforderungen wie Betriebskontinuität und Datenschutz. Optische Intel® Silizium-Photonik-Transceiver ermöglichen großen Rechenzentren den Einsatz von 100-Gbps-Lösungen, um Switches über hunderte Meter oder mehrere Kilometer anstatt über nur wenige Meter zu verbinden. Der einzigartige Ansatz von Intel integriert hybride Silizium-Laser als Teil des Photonik-Chips. Das ermöglicht die Herstellung und Prüfung auf Wafern und sorgt für branchenführende Qualität bei den hohen Volumen, die Hyperscaler benötigen, um ihren wachsenden Bedarf an Daten zu befriedigen. Da sich die Bandbreite innerhalb der Switching-Infrastruktur in diesen riesigen Rechenzentren von Switches mit 3,2/6,4 Tbps auf Switches mit 12,8 und 25,6 Tbps erhöht hat, wächst auch der Bedarf an Glasfaserverbindungen mit höherer Bandbreite. Intel bleibt weiterhin Innovationstreiber, indem es auch optische Intel® Silizium-Photonik Module mit 200 Gbps und 400 Gbps herstellt. Große private Rechenzentren sollten in Erwägung ziehen, dem Beispiel der Hyperscaler zu folgen und diese ausgereifte und kostengünstige Technik einzusetzen.

## Sicherheit und Schutz für Daten, Anwendungen und Zusammenarbeit

Je umfangreicher eine Infrastruktur wird, desto größer werden die Sicherheitsbedrohungen und Geschäftsrisiken. Datenschutz ist ein Hauptanliegen für Unternehmen, die die Public Cloud nutzen, und hat somit auch für CSPs sehr hohe Priorität. [Confidential Computing](#) ist eine aufstrebende Brancheninitiative, die sich auf die sichere Verarbeitung von Daten konzentriert. Es ermöglicht, verschlüsselte Daten im Arbeitsspeicher zu verarbeiten, ohne sie dem Rest des Systems zugänglich zu machen. So wird die Gefährdung von sensiblen Daten verringert und die Nutzer erhalten mehr Kontrolle und Transparenz. Intel ist Gründungsmitglied des Confidential Computing Consortium und trägt mit seinen Techniklösungen dazu bei, dass CSPs – und sonstige Anbieter – in ihren Rechenzentren für Daten- und Anwendungsschutz sorgen können.

### Intel® Software Guard Extensions (Intel® SGX)

Die Anforderungen an den Datenschutz steigen ständig. Sogar Daten, die während der Speicherung und der Übertragung durch Verschlüsselung gut geschützt sind, können während der Verarbeitung gefährdet sein. CSPs und andere Rechenzentren nutzen [Intel® Software Guard Extensions](#) (Intel® SGX), um Daten in diesem kritischen Moment der Verarbeitung, wenn sie nicht verschlüsselt sind, zu schützen. Das erhöht die Sicherheit bei Anwendungsfällen, wie zum Beispiel:

- **Föderiertes (maschinelles) Lernen**. Föderiertes Lernen (Federated Learning, FL) ist ein verteilter Ansatz von maschinellem Lernen (ML), der es mehreren Unternehmen bzw. Organisationen ermöglicht, gemeinsam an ML-Projekten zu arbeiten, jedoch erfordert, dass sensible Daten geschützt werden. Intel® SGX ist ideal zum Aufbau von hardwarebasierten vertrauenswürdigen Ausführungsumgebungen für Federated-Learning-Lösungen in der Cloud.
- **Vertrauliche Container und VMs**. In mandantenfähigen Cloud-Umgebungen sorgen sich Kunden, dass Container und VMs angreifbar sein könnten. Dass Intel® SGX vertrauenswürdige Ausführung unterstützt lässt sich dazu nutzen, Container- und VM-Prozesse vor Angriffen von außen zu schützen.
- **Vertrauliche Datenbanken**. Viele Unternehmen und Organisationen verlagern Datenbanken in die Public Cloud. Intel® SGX kann dazu genutzt werden, den Schutz solcher Datenbanken durch Isolierung sensibler Daten oder kryptografischer Schlüssel zu verbessern.
- **Blockchains**. Intel® SGX hilft CSPs, bei Blockchains den Schutz und die Sicherheit für die Transaktionsverarbeitung, das Konsensverfahren, die Smart Contracts und die Schlüsselspeicherung zu verbessern.

Intel® SGX ist ein Befehlssatz, der der Erstellung sicherer Enklaven dient – kleine, vertrauenswürdige Umgebungen innerhalb einer CPU, in denen Code so ausgeführt werden kann, dass er für das normale Betriebssystem nicht zugänglich ist. Eine Attestierung dieser Enklaven ist auch aus der Ferne möglich, wodurch eine Partei kryptografisch verifizieren kann, dass eine auf dem Computer einer anderen Partei laufende Enklave vertrauenswürdigen, unveränderten Code ausführt.

Intel® SGX bietet der Cloud eine zusätzliche Sicherheitsebene. CSPs haben nun damit begonnen, Kunden die Möglichkeit anzubieten, ihre sensiblen Daten so vertraulich zu behandeln, dass selbst der CSP nicht darauf zugreifen kann.<sup>3</sup> Auch im Fall von unerkannter Malware oder eines bösartigen Administrators schützt Intel® SGX die Daten des Kunden vor Offenlegung.

Umfragen zeigen, dass die größte Sorge von Cyber-Security-Experten hinsichtlich Cloud-Sicherheit ein Datenverlust oder Datenleck ist.<sup>4</sup> Und 70 Prozent der 2020 für eine Studie befragten Führungskräfte waren der Meinung, dass Sicherheitsbedenken einer umfassenderen Nutzung der Public Cloud im Wege standen.<sup>5</sup> CSPs sehen in dem höheren Maß an Sicherheit und Schutz, das Intel® SGX bietet, eine große Chance, die Bedenken ihrer Kunden hinsichtlich des Schutzes der Vertraulichkeit von Daten in der Public Cloud zu zerstreuen. Das könnte die Migration in die Cloud in noch größerem Umfang fördern.

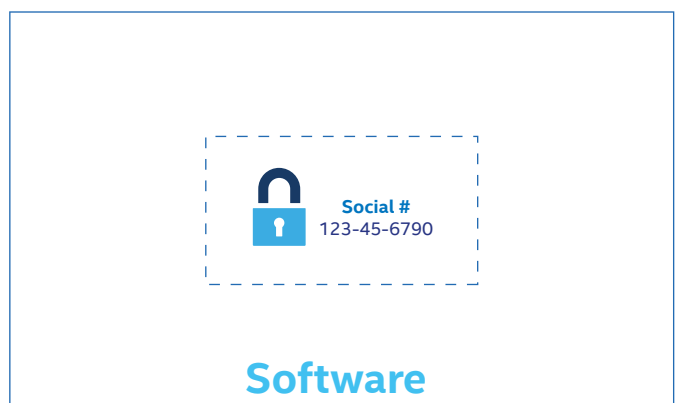
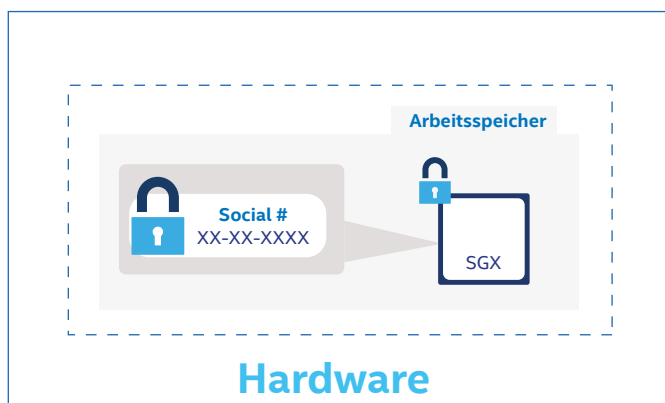


Abbildung 1: Intel® SGX schützt Informationen in einer Enklave, wenn die Anwendungen läuft, und in der Hardware, wenn sie nicht läuft

## Intel® QuickAssist-Technik (Intel® QAT)

Die [Intel® QuickAssist-Technik \(Intel® QAT\)](#) bietet Hardwarebeschleunigung für rechenintensive Operationen wie Chiffrierungsverfahren, Hashfunktionen, Verschlüsselungen mit öffentlichen Schlüsseln und Kompression. Indem die CPU-Kerne von diesen Aufgaben entlastet werden, ist durch den Einsatz von Intel® QAT eine Performancesteigerung um das bis zu 4,3-fache möglich.<sup>6</sup>

Intel® QAT bietet außerdem eine wichtige Sicherheitsfunktion namens Intel® Key Protection-Technik (Intel® KPT), die private kryptografische Schlüssel schützt. <sup>7</sup> Durch Intel® KPT wird ein privater Schlüssel verschlüsselt, bevor er in die VM gelangt, die ihn benutzen muss. Nur Intel® QAT innerhalb des Chips kann den privaten Schlüssel entschlüsseln, wodurch dieser auf der Hardware-Ebene besser geschützt ist.

CSPs profitieren von den Beschleunigungsfunktionen von Intel® QAT, da dadurch die Anzahl der für die Kunden verfügbaren VMs erhöht wird. Sie können die Funktionalität von Intel® KPT nutzen, um ihre Sicherheit zu differenzieren. Dadurch können immer raffiniertere Software- und Hardware-Angriffen bekämpft werden und erfüllen ihre Clouds regulatorische Anforderungen.

## KI und HPC

Rechenintensive Workloads – KI, Advanced Analytics und Simulationen – werden von Cloud-Diensten immer besser unterstützt. Cloud-Computing versetzt Unternehmen und Organisationen in die Lage, ihre Computing-Ressourcen je nach Anwendungsanforderungen zu erhöhen oder zu verringern. Hyperscaler stellen in Zusammenarbeit mit Intel eine Infrastruktur und Ressourcen zur Verfügung, die ideal für rechenintensive Anwendungen sind, die über einen längeren Zeitraum viel Rechenleistung benötigen. Das datenzentrierte Portfolio von Intel geht auf die Herausforderungen rechenintensiver Workloads ein, indem HPC, KI-Beschleunigung und Advanced Data Analytics in einer einzelnen Computing-Umgebung vereint werden, die CSPs dazu nutzen können, ihre Kunden bei wissenschaftlichen Simulationen, Finanzanalytik, KI/Deep Learning (DL) sowie 3D-Modellierung und -Analyse zu unterstützen.<sup>8</sup>

In den folgenden Abschnitten werden einige der Intel Lösungen beschrieben, die HPC- und KI-Anwendungen in der Cloud ermöglichen.

## Intel® Advanced Vector Extensions 512 (Intel® AVX-512)

Bei der Vektorverarbeitung wird in einem Array von Integer- oder Gleitkommazahlen gleichzeitig eine arithmetische Rechenoperation durchgeführt, die bei Anwendungen wie wissenschaftlichen Simulationen und 3D-Modellierung sehr aufwändig sein kann.

[Intel® Advanced Vector Extensions 512](#) (Intel® AVX-512) ist ein CPU-Befehlssatz, der die rechenintensive Vektorverarbeitung von Compute-Workloads um das bis zu 1,6-fache steigert.<sup>9</sup>

Intel® AVX-512 hat Auswirkungen auf Rechen-, Speicher- und Netzwerkfunktionen. Die Zahl 512 bezieht sich auf die Breite der Registerdatei (in Bit), die die Parameter für die Menge von Daten festlegt, mit der ein Befehlssatz gleichzeitig operieren kann. Intel® AVX-512 ermöglicht im Vergleich zu seinem Vorgänger Intel® AVX2 doppelt so viele Gleitkommaoperationen pro Sekunde (FLOPS). Intel® AVX-512 ermöglicht es also, mit einem einzelnen Befehl doppelt so viele Datenelemente wie Intel® AVX2 und vier Mal so viele wie Intel® Streaming SIMD Extensions (Intel® SSE) zu verarbeiten.

Intel arbeitet gemeinsam mit Hyperscalern an der Beschleunigung von Operationen ohne Code-Modifikationen an ihren bestehenden Frameworks.

Den größten Performance-Gewinn zeigt Intel® AVX-512 bei Workloads, die die gleiche Art von Vektor/Matrix-Operationen an großen Datenmengen erfordern wie zum Beispiel DNA-Sequenzierung.<sup>10</sup> Dank der Fähigkeit, mit mehr Informationen gleichzeitig zu arbeiten, hilft Intel® AVX-512 dabei, Rechenaufgaben zu bewältigen und die Performance für Workloads und Anwendungen wie KI/DL, wissenschaftliche Simulationen, Finanzanalytik sowie 3D-Modellierung und -Analyse zu beschleunigen.

Wenn man diese Art von KI- und HPC-Workloads in eine Public Cloud verlagern möchte, sollte man sich unbedingt für eine Instanz von einem CSP entscheiden, der Intel® AVX-512 anbietet, um so ein hervorragendes Preis-/Leistungsverhältnis zu erhalten.

## Intel® Deep Learning Boost (Intel® DL Boost)

DL-Anwendungen erfordern HPC-Fähigkeit und eine niedrige Latenz. Traditionell wurden Grafikprozessoren (GPUs) für die Ausführung von ML- und DL-Workloads eingesetzt. Dadurch kam es zu höheren Hardwarekosten, wenn man die für bestimmte Cloud-Instanzen erforderliche Performance erreichen wollte. Fortschritte bei der CPU-Technologie haben CSPs neue Möglichkeiten eröffnet, ihre zentralen Infrastrukturservices um integrierte KI-Beschleunigung zu erweitern.

CSPs setzen [Intel® Deep Learning Boost](#) (Intel® DL Boost) in leistungsstarken Instanzen ein, um Kunden die optimale Umgebung für innere Schleifen von Convolutional Neural Networks und andere rechenintensive Workloads zu bieten. Mit Intel® DL Boost lässt sich bei dieser Art von KI-Operationen die Performance um das bis zu 3,4-fache steigern.<sup>11</sup>

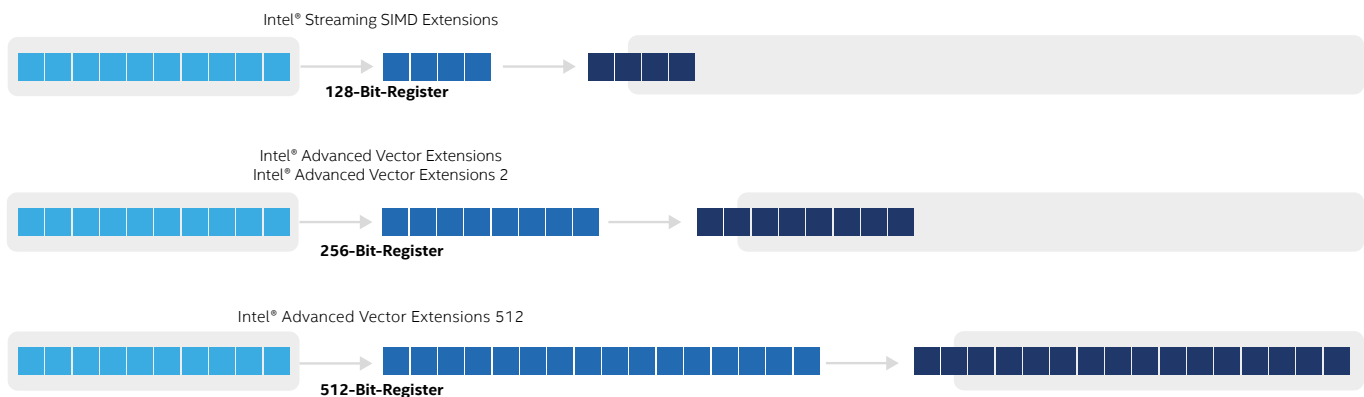


Abbildung 2: Intel® AVX-512 verarbeitet mehr Daten mit einem einzelnen Befehl

Intel® DL Boost bezeichnet eine Reihe von Intel® AVX-512 Befehlen namens Vector Neural Network Instructions (VNNI), die die Basis von Intel® AVX-512 durch vier neue Befehle zur Beschleunigung innerer Schleifen von Convolutional Neural Networks erweitern.<sup>12</sup> Intel® DL Boost kann die Performance dieser Art von ML, die bei KI-Anwendungen wie Bilderkennung, Videoanalyse und natürlicher Sprachverarbeitung (NLP) zum Einsatz kommt, deutlich steigern.<sup>13</sup>

Einfach durch Aktivierung von Intel® DL Boost kann dieselbe Hardware-Plattform, die Kunden für andere HPC-Workloads verwenden, auch für KI-Workloads optimiert werden.

KI/DL-Initiativen profitieren von Intel® DL Boost – egal ob auf derselben lokalen HPC-Infrastruktur, die für andere Workloads genutzt wird, oder in der Cloud durch Nutzung einer Instanz eines Hyperscalers, der dieselbe Funktionalität bietet.

### Intel® Field Programmable Gate Arrays (FPGAs)

Abgesehen von den bereits erwähnten SmartNIC-Anwendungsfällen von [Intel® FPGAs](#) werden Intel® FPGAs durch das harte Codieren von Operationen in die Hardware auch zur Beschleunigung vieler anderer Workloads wie zum Beispiel ML-Operationen eingesetzt. Zu den Anwendungen, die stark von ML-Implementierungen auf einem FPGA profitieren, gehören u. a. Intelligent Vision, wissenschaftliche Simulationen, Biowissenschaften sowie medizinische Datenanalyse.<sup>14</sup>

Da sie sich für verschiedene Typen von ML-Modellen rekonfigurieren lassen, können Intel® FPGAs zur Beschleunigung von KI-Operationen genutzt werden, bei denen es mehrmals im Jahr zu größeren algorithmischen Änderungen kommt. Intel® FPGAs ermöglichen niedrige Latenzen für Echtzeit-Inferenzanfragen.<sup>15</sup> Implementierungen von Intel® FPGA Neuralprozessoren (Neural Processing Units, NPU) benötigen kein Batching, wodurch die Latenz viel geringer als bei CPUs oder GPUs ist.

CSPs stellen ihren Kunden Beschleunigung per Intel® FPGAs sowohl in der Cloud als auch am Edge zur Verfügung.<sup>16</sup> Diese zusätzliche Beschleunigung für Echtzeit-Inferenzanfragen macht die Migration rechenintensiver Workloads von lokalen HPC-Umgebungen in die Cloud viel attraktiver.

### Intel® Math Kernel Library for Deep Neural Networks (Intel® MKL-DNN)

Die [Intel® Math Kernel Library for Deep Neural Networks](#) (Intel® MKL-DNN) beschleunigt DL-Frameworks auf Intel® Architektur. Sie umfasst hochgradig vektorisierte und Thread-basierte Bausteine für die Implementierung von Convolutional Neural Networks mit Schnittstellen für C und C++. Intel arbeitet gemeinsam mit Hyperscalern an der Integration von Intel® MKL-DNN in die von ihnen genutzten Frameworks wie Apache MXNet\* und TensorFlow\*.

### Intel® OpenVINO™ Toolkit

Die Intel® Distribution des OpenVINO™ (Open Visual Inference and Neural Network Optimization) Toolkits ermöglicht Entwicklern den schnellen Einsatz von Anwendungen und Lösungen, die das menschliche Sehvermögen nachahmen. Sie beschleunigt auch andere KI-Workloads wie Audio, Sprachausgabe, Sprache und Empfehlungssysteme.

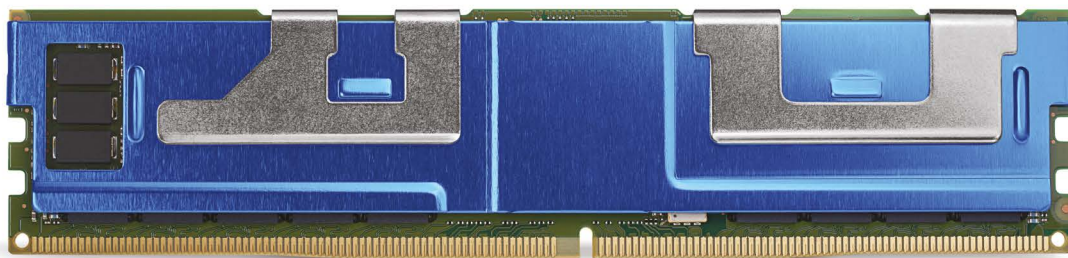
Der [Intel® OpenVINO™ Toolkit](#) ermöglicht DL-Inferenz am Edge auf heterogener Intel® Hardware wie:

- Intel® CPUs
- Integrierte Intel® Grafik
- Intel® FPGAs
- Intel® Movidius™ Neural Compute Stick (NCS) oder NCS 2
- Intel® Vision Accelerator Design mit Intel® Movidius™ Vision Processing Units (VPUs)

Der Intel® OpenVINO™ Toolkit kann zur Implementierung von Computer-Vision-Inferenz auf Edge-Geräten eingesetzt werden, wobei Modelle genutzt werden, die auf von CSPs bereitgestellten leistungsstarken Instanzen trainiert wurden.<sup>17</sup>

### Steigerung der Workload-Performance

Anwendungen sind entscheidend, um heutzutage konkurrenzfähig zu sein, und ihre Performance hängt davon ab, dass die richtigen Rechenzentrumsressourcen zum richtigen Zeitpunkt zur Verfügung stehen. Ob revolutionär und neu wie persistenter Intel® Optane™ Speicher oder bewährt und grundlegend wie die Intel® Turbo-Boost-Technik: CSPs und ihre Kunden verlassen sich auf Intel, wenn es um Performancesteigerung auf jeder Ebene geht.



### Persistenter Intel® Optane™ Speicher

  
**Groß**  
Hohe Kapazität  
für Skalierbarkeit

  
**Persistent**  
Datenpersistenz ohne  
Disk-I/O (Input/Output)

  
**Preisgünstig**  
Geringe Kosten

  
**Robust**  
Weniger Ausfallzeiten

Abbildung 3: Vorteile von persistentem Intel® Optane™ Speicher

## Persistenter Intel® Optane™ Speicher

[Persistenter Intel® Optane™ Speicher](#) ist eine innovative Klasse von Arbeits- und Datenspeichertechnik, die es CSPs ermöglicht, dass ihre Kunden größere Datenmengen näher am Prozessor speichern – bei niedriger Latenz und einer konstanten DRAM-ähnlichen Performance.

CSPs nutzen persistenten Intel® Optane™ Speicher zur kostengünstigen Erweiterung der Arbeitsspeicherkapazität, um größere Mengen an „heißen“ Daten für die Verarbeitung durch anspruchsvolle Workloads zu ermöglichen. In-Memory-Datenbanken sind ein Paradebeispiel für Workloads, die von einer Skalierung des Arbeitsspeichers profitieren, da sie am effizientesten arbeiten, wenn eine In-Memory-Verarbeitung des gesamten Datenbestands möglich ist. Die hohen Kosten von DRAM drücken die Kosteneffizienz einer Skalierung des Arbeitsspeichers.

SAP HANA\* ist Vorreiter bei der Optimierung von In-Memory-Datenbankmanagementsystemen (DBMS) für persistenten Intel® Optane™ Speicher – einschließlich der erforderlichen Modifikationen, um dessen Persistenz in vollem Umfang nutzen zu können. CSPs haben daher daran gearbeitet, Angebote mit Instanzen einzuführen, die dafür optimiert sind, SAP HANA\* in der Cloud auf skalierbaren Intel® Xeon® Prozessoren mit großen Mengen an persistentem Intel® Optane™ Speicher laufen zu lassen.<sup>18</sup> Unternehmen, die SAP-Lösungen lokal laufen lassen, profitieren auch davon, ihre privaten Clouds mit persistentem Intel® Optane™ Speicher aufzurüsten. So können sie die Vorteile der Speicherpersistenz nutzen, wozu beispielsweise die um Größenordnungen schnelleren Neustartzeiten gehören.<sup>19</sup>

CSPs finden persistenten Intel® Optane™ Speicher auch für den Betrieb ihrer eigenen riesigen Inhouse-Datenbanksysteme nützlich.<sup>20</sup>

Die Intel® Optane™ Technik füllt die Lücke zwischen teurerem DRAM als Arbeitsspeicher und langsamerem NAND als Datenspeicher. Persistenter Speicher, der als DIMM am Speicherbus angeschlossen wird, stellt bei der Skalierung des Arbeitsspeichers eine kostengünstige Alternative zu DRAM dar – mit den zusätzlichen Vorteilen der Speicherpersistenz, die immer stärkere Verbreitung finden werden, wenn mehr Softwarehersteller entsprechende Optimierungen vornehmen.

## Intel® Optane™ Solid-State-Laufwerke (SSDs)

Dieselbe Intel® Optane™ Technik, die größeren und erschwinglicheren Arbeitsspeicher unterstützt, kommt auch bei [Intel® Optane™ Solid-State-Laufwerken \(SSDs\)](#) zum Einsatz, wo sie im Datenspeicher-Tier für hohe Performance und hohe Haltbarkeit sorgt. Während Intel® QLC 3D NAND-SSDs den Bedarf von CSPs an massiver Speicherkapazität decken, sind Intel® Optane™ SSDs für Szenarien interessant, in denen eine geringere Latenz und eine größere Ausdauer erforderlich sind, wie zum Beispiel im Caching-Tier.

CSPs ist klar, wie nützlich Intel® Optane™ SSDs für das Caching sind, da diese dazu fähig sind, große Datenmengen schnell zu lesen und zu schreiben. Wenn Kunden-Workloads mit immer größeren Datenmengen in die Cloud migriert werden, kann sich die Caching-Geschwindigkeit zu einem Engpass entwickeln, der zum Beispiel komplexe Analyse- und KI-Anwendungen daran hindert, so schnell wie möglich Erkenntnisse aus den Daten abzuleiten. Die Skalierung eines schnelleren Caching-Tiers steigert die Performance dieser komplexen Workloads und trägt dazu bei, die Kundenerwartungen besser zu erfüllen.

Ein weiteres Merkmal der Intel® Optane™ SSDs, das für CSPs wichtig ist, ist ihre Haltbarkeit. Wie Autoreifen sind SSDs für eine bestimmte „Laufleistung“ ausgelegt, die sie erreichen, bevor sie abgenutzt sind. Im Falle von SSDs wird diese Haltbarkeit in während des Lebenszyklus des Laufwerks insgesamt geschriebenen Petabytes (PBW) gemessen. Hyperscaler können Cache-Laufwerke ohne weiteres 24 Stunden am Tag an die Grenze ihrer Lese-/Schreibgeschwindigkeit bringen. Deshalb ist die

Haltbarkeit ein kritischer Faktor bei den Gesamtbetriebskosten (TCO) für diese Laufwerke. Ein Hyperscaler stellte fest, dass zwei Intel® Optane™ SSDs in einem Array mehr als vier Mal so viele PBW als ein Array mit sechs NAND-SSDs schaffen und gleichzeitig beim Caching für eine geringere Latenz sorgen.<sup>21</sup>

Intel® Optane™ SSDs unterstützen CSPs – und private Clouds – bei der Skalierung eines schnellen Caching-Tiers, der riesige Datenmengen schreibt, ohne die Laufwerke dabei so schnell zu verschleifen.

## Intel® Hyper-Threading-Technik (Intel® HT-Technik)

Einige etablierte Lösungen, die ursprünglich von Intel entwickelt wurden, spielen für CSPs nach wie vor eine zentrale Rolle bei der Bereitstellung von Services in großem Maßstab. Die [Intel® Hyper-Threading-Technik](#) ist ein gutes Beispiel dafür.

Die Intel® HT-Technik wird eingesetzt, um die parallele Verarbeitung zu verbessern, wodurch die Anzahl der unabhängigen Befehle in der Pipeline erhöht wird. Mit der Intel® HT-Technik fungiert ein physischer Kern für das Betriebssystem als zwei Prozessoren und ermöglicht so das parallele Scheduling von zwei Prozessen pro Kern. CSPs nutzen die Intel® HT-Technik normalerweise dazu, Kunden doppelt so viele virtuelle CPUs anzubieten, wie es physische Kerne gibt. Die Threads eines Kerns werden dabei niemals zwischen zwei Kunden aufgeteilt, um Sicherheitsrisiken aufgrund gemeinsam genutzter Ressourcen zu vermeiden. CSPs bieten Kunden auch die Möglichkeit, die Intel® HT-Technik für solche Anwendungen zu deaktivieren, die eine hohe Performance aus einem einzelnen Thread pro Kern benötigen.

## Intel® Turbo-Boost-Technik

Die [Intel® Turbo-Boost-Technik](#) ist eine weitere bewährte Lösung, die in On-Premises-Umgebungen wohl als selbstverständlich angesehen wird. Bei der Verlagerung von Workloads in die Cloud sollte sichergestellt werden, dass auf Intel® Prozessoren basierende Instanzen zum Einsatz kommen, damit die gewünschte hohe Performance erreicht wird.

Die Intel® Turbo-Boost-Technik steigert die Prozessor-Performance bei Lastspitzen. Sie lässt Prozessorkerne automatisch mit einer höheren als der nominalen Taktfrequenz laufen, wenn sie unterhalb der Spezifikationsgrenzen für Leistung, Strom und Temperatur arbeiten. Dadurch können CSPs ihren Kunden mehr Performance bieten, wenn deren Workloads dies erfordern. Das ist einer der Gründe dafür, dass CSPs höhere Preise für Intel® basierte Instanzen verlangen können.

## Großartige Clouds bauen auf Intel® Technik

Hyperskalierung stellt die Top-CSPs vor die gleichen Herausforderungen wie typische Rechenzentren. Hyperscaler sind die ersten, die auf neue Probleme und neue Grenzen stoßen, wenn sie ihren Geschäftsbetrieb in bisher ungekanntem Ausmaß ausbauen. Und wie dieser Artikel gezeigt hat, vertrauen CSPs auf Intel® Technik, um Probleme zu lösen und die Grenzen der Hyperskalierung heute und in Zukunft zu überwinden.

Von der Erfahrung der Hyperscaler kann man profitieren. Kunden von Cloud-Diensten hilft das Wissen über die den Cloud-Instanzen zugrundeliegende Technologie bei der Auswahl der richtigen Service-Optionen. Dadurch können sie sicherstellen, dass die von ihnen benötigten Funktionen enthalten sind und eine nahtlose Migration gewährleistet wird. Betreiber einer privaten Cloud erhalten durch Kenntnis der den Top-CSPs zugrundeliegenden Technologie ein Best-Practice-Beispiel. Dieses hilft ihnen bei der Auswahl der besten verfügbaren Technologien zur Bewältigung der Herausforderungen ihres eigenen Rechenzentrums. Durch Anwendung der Best Practices der Hyperscaler können sie die Markteinführungszeit verkürzen und die private Cloud zuverlässig skalieren.

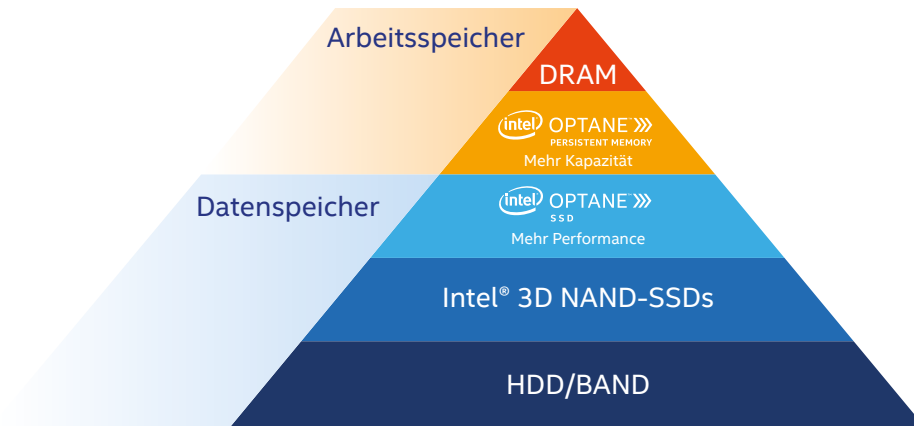


Abbildung 4: Die Intel® Optane™ Technik füllt die Lücke zwischen teurerem Arbeitsspeicher und langsamem Datenspeicher

Unternehmen entwickeln ihre Geschäftsstrategien weiter und setzen Cloud Computing-Services weiterhin ein. Deshalb hilft tieferes technisches Wissen über die Cloud dabei, Risiken zu minimieren und Leistung, Performance und Auswahl zu bieten.

Eine Cloud-Strategie, die auf Intel® Architektur setzt, erhöht auch die Agilität und Flexibilität bei der Auswahl des CSP und der Architektur, die den Anforderungen des Unternehmens am besten entsprechen.

Für Intel® Architektur optimierte Workloads helfen Unternehmen, Daten und Anwendungen einfacher zwischen verschiedenen Clouds zu verschieben, die dieselben wegweisenden Intel® Techniklösungen unterstützen. Das nutzt besonders der wachsenden Zahl von Unternehmen, die lokale Rechenzentren betreiben und auch mehrere Public Clouds nutzen. Dieselben Best Practices, die Hyperscaler nutzen, lassen sich auch in der privaten Cloud anwenden. So können Workloads an mehreren Standorten ausgeführt werden – je nachdem, wie es das Geschäft oder die Sicherheit erfordern.

## Weitere Informationen

Solution Brief zu Amazon Web Services\* (AWS\*): [intel.de/content/www/de/de/cloud-computing/amazon-ec2-cloud-instances-brief.html](https://www.intel.de/content/www/de/de/cloud-computing/amazon-ec2-cloud-instances-brief.html)

Webpage zu AWS\* und Intel: [intel.de/content/www/de/de/big-data/partners/aws.html](https://www.intel.de/content/www/de/de/big-data/partners/aws.html)

Solution Brief zu Azure\*: [intel.de/content/www/de/de/cloud-computing/azure-stack-and-intel-drive-business-value-brief.html](https://www.intel.de/content/www/de/de/cloud-computing/azure-stack-and-intel-drive-business-value-brief.html)

Webpage zu Azure\* und Intel: [intel.de/content/www/de/de/big-data/partners/microsoft/overview.html](https://www.intel.de/content/www/de/de/big-data/partners/microsoft/overview.html)

Solution Brief zu Google Cloud\*: [intel.de/content/www/de/de/cloud-computing/google-cloud-brief.html](https://www.intel.de/content/www/de/de/cloud-computing/google-cloud-brief.html)

Webpage zu Google Cloud\* und Intel: [intel.de/content/www/de/de/big-data/partners/google-cloud-platform.html](https://www.intel.de/content/www/de/de/big-data/partners/google-cloud-platform.html)





- <sup>1</sup> Basierend auf Aufzeichnungen von Intel IT aus den ersten beiden Wochen von COVID-19.
- <sup>2</sup> Laut Dave Maltz, Distinguished Engineer bei Microsoft in der Abteilung Azure Networking, setzen mehr als 10 Hyperscaler und Cloud-Services SONIC\* als Switching-Betriebssystem ein, wobei Microsoft und Alibaba die beiden größten CSPs sind. Quelle: The Next Platform. „Is Microsoft's SONIC Winning the War of the Noses?“, Mai 2020. [nextplatform.com/2020/05/12/is-microsofts-sonic-winning-the-war-of-the-noses/](https://nextplatform.com/2020/05/12/is-microsofts-sonic-winning-the-war-of-the-noses/).
- <sup>3</sup> Microsoft Azure\* kündigte im April 2020 als erster CSP die allgemeine Verfügbarkeit von Confidential Computing auf der Basis von Intel® SGX an. Quelle: Microsoft „DCsv2-Series VM Now Generally Available from Azure Confidential Computing“, April 2020. <https://azure.microsoft.com/de-de/blog/dcsv2series-vm-now-generally-available-from-azure-confidential-computing/>. Nutzung mit Genehmigung von Microsoft.
- <sup>4</sup> In einer Umfrage unter 700 IT- und Sicherheitsfachleuten, gaben 81 Prozent der Cloud-Nutzer an, dass sie erhebliche Sicherheitsbedenken haben, insbesondere Bedenken hinsichtlich der Risiken von Datenverlusten und -lecks (62 Prozent), dicht gefolgt von Bedenken hinsichtlich der Einhaltung gesetzlicher Vorschriften (57 Prozent). Quelle: AlgoSec. „Cloud Security Alliance Study Identifies New and Unique Security Challenges in Native Cloud, Hybrid and Multi-cloud Environments“, Mai 2019. [globenewswire.com/news-release/2019/05/21/1833639/0/en/Cloud-Security-Alliance-Study-Identifies-New-and-Unique-Security-Challenges-in-Native-Cloud-Hybrid-and-Multi-cloud-Environments.html](https://globenewswire.com/news-release/2019/05/21/1833639/0/en/Cloud-Security-Alliance-Study-Identifies-New-and-Unique-Security-Challenges-in-Native-Cloud-Hybrid-and-Multi-cloud-Environments.html). Siehe auch: Cybersecurity Insiders. „2019 Cloud Security Report (ISC)2“. [cybersecurity-insiders.com/portfolio/2019-cloud-security-report-isc2/](https://cybersecurity-insiders.com/portfolio/2019-cloud-security-report-isc2/).
- <sup>5</sup> Barracuda Networks. „Future Shock: The Cloud Is the New Network“, März 2020. [https://lp.barracuda.com/BEU-AMER-WBN-20200304-FutureShockCloudReport\\_LP-Registration.html](https://lp.barracuda.com/BEU-AMER-WBN-20200304-FutureShockCloudReport_LP-Registration.html).
- <sup>6</sup> 4,3-fache Performancesteigerung mit Intel® QAT, gemessen mit einem Private Key Exchange TLS 1.2 RSA2K Workload am 5. November 2019: Intel® Server-Mainboard S2600WFD mit Intel® Xeon® Gold 6252N Prozessor (2,3 GHz, 24 Kerne, 2 UPI-Verbindungen) bei dem 18 Kerne/36 Threads mit Intel® Turbo-Boost-Technik liefern, mit Intel® QuickAssist Adapter 8970, 12 x DDR4 @ 2.933 MHz mit 32 GB, BIOS: SE5C620. 86B.0X.02.0040.060420190144; Microcode: 0x5000026, Ubuntu\* 19.04, 5.0.0-23-generic, GCC\* 8.3 Compiler, Intel® Ethernet-Netzwerkkardadapter XXV710-DA2, NGINX\* 1.14.2, OpenSSL\* 1.1.0k, Intel® QAT Engine v0.54.1, Intel® QAT Treiber L05000007.
- <sup>7</sup> Intel. „Intel® QuickAssist-Technik mit Intel® Key Protection-Technik in Intel® Server-Plattformen auf der Basis skalierbarer Intel® Xeon® Prozessoren“, 2017. [intel.de/content/www/de/de/architecture-and-technology/key-protection-technology-white-paper.html](https://intel.de/content/www/de/de/architecture-and-technology/key-protection-technology-white-paper.html).
- <sup>8</sup> Für weitere Informationen über das datenzentrierte Portfolio von Intel siehe: Intel. „Intel's Data-Centric Portfolio Accelerates Convergence of High-Performance Computing and AI Workloads“, Juni 2019. <https://newsroom.intel.com/news/intel-data-centric-portfolio-accelerates-convergence-high-performance-computing-ai-workloads/>.
- <sup>9</sup> Durchschnittlich 1,6-fache Performancesteigerung mit Intel® AVX-512, gemessen mit Finanzdienstleistungs-Workloads am 1. November 2019: Intel® Server-Mainboard S2600WFD mit 2-Sockel Intel® Xeon® Platinum 8268 Prozessoren (2,9 GHz, 24 Kerne, 2 UPI-Verbindungen), 12 x DDR4 @ 2.933 MHz mit 16 GB, eine SSD, BIOS: SE5C620.86B.02.01.0008.031920191559; Microcode: 0x500001c, Red Hat Enterprise Linux\* 7.7, Kernel 3.10.0-1062.1.1. FSI-Kernel v2.0: Geometrisches Mittel (3 Workloads: Binomialmodell für Optionspreise, Black-Scholes-Modell, Monte-Carlo-Modell), Intel® AVX2 256 Build im Vergleich zu Intel® AVX-512 Build, Intel® Compiler 2019u5, Intel® Math Kernel Library (Intel® MKL) 2019u5, BIOS: Binomialmodell (Intel® HT-Technik aktiviert, Intel® Turbo-Boost-Technik aktiviert, SNC deaktiviert, 3 Threads/Kern), Black-Scholes-Modell (Intel® HT-Technik deaktiviert, Intel® Turbo-Boost-Technik aktiviert, SNC deaktiviert, 1 Thread/Kern), Monte-Carlo-Modell (Intel® HT-Technik aktiviert, Intel® Turbo-Boost-Technik aktiviert, SNC deaktiviert, 2 Threads/Kern).
- <sup>10</sup> Siehe beispielsweise: DeepVariant Blog. „The Power of Building on an Accelerating Platform: How DeepVariant Uses Intel's AVX-512 Optimizations“, April 2019. <https://google.github.io/deepvariant/posts/2019-04-30-the-power-of-building-on-an-accelerating-platform-how-deepvariant-uses-intels-avx-512-optimizations/>.
- <sup>11</sup> 3,4-fache Performancesteigerung mit Intel® DL Boost (VNNI) gemessen mit ResNet-50\* Inferenzdurchsatzleistung; getestet von Intel am 25.12.2019: 1 Knoten, 2 x Intel® Xeon® Gold 6258R Prozessor auf Intel® Referenzplatte mit insgesamt 384 GB Arbeitsspeicher (12 Steckplätze; 32 GB, 2.933 MHz), uCode: 0x500002c, Intel® HT-Technik aktiviert, Intel® Turbo-Boost-Technik aktiviert, mit Ubuntu\* 19.1, 05.3.0-24-generic, AIXPRT\* Bilderkennung AIXPRT v1.01, Intel® Distribution des OpenVINO™ Toolkits Version 2019 R3, ResNet50\* v1, für INT8 mit Intel® DL Boost: BS=4, 56 Instanzen, für FP32 BS=4, 56 Instanzen.
- <sup>12</sup> Intel. „Lower Numerical Precision Deep Learning Inference and Training“, Januar 2018. [intel.com/content/www/us/en/artificial-intelligence/solutions/lower-numerical-precision-deep-learning-inference-and-training.html](https://intel.com/content/www/us/en/artificial-intelligence/solutions/lower-numerical-precision-deep-learning-inference-and-training.html).
- <sup>13</sup> Bei Tests, bei denen die Inferenz von Bildern/Sekunde pro Sockel mittels PyTorch\* gemessen wurde, verbesserte die Integration von Intel® MKL-DNN die Performance in verschiedenen Szenarien von FP32 und INT8 um das 7,7- bis 105-fache des Ausgangswerts (FP32 ohne Intel® MKL-DNN) für ResNet50\*, Faster R-CNN\* und RetinaNet bei einer Batch-Größe von 1 auf einem Intel® Xeon® Platinum 8280 (Cascade Lake) Prozessor mit einem Sockel. Quelle: Intel und Facebook\* Collaborate to Boost PyTorch\* CPU Performance“, April 2019. <https://software.intel.com/content/www/us/en/development/articles/intel-and-facebook-collaborate-to-boost-pytorch-cpu-performance.html>.
- <sup>14</sup> Siehe zum Beispiel das laufende Microsoft/Projekt Brainwave. 2018 führten Bing\* und Microsoft Azure\* neue Multi-FPGA-Appliances in Rechenzentren ein, die das Verhältnis der Rechenleistung zwischen CPUs und FPGAs verschoben. Dabei wurde jeder Server mit mehreren Intel® Arria\* 10 FPGAs ausgestattet. Quelle: Microsoft. „Project Catapult“. [microsoft.com/en-us/research/project/project-catapult/](https://microsoft.com/en-us/research/project/project-catapult/). Nutzung mit Genehmigung von Microsoft.
- <sup>15</sup> Der Intel® Stratix\* 10 NX FPGA ermöglicht beispielsweise einen bis zu 15-fachen INT8-Durchsatz für KI-Anwendungen. Auf Basis interner Schätzwerte von Intel. Für weitere Informationen siehe: Intel. „Intel® Stratix\* 10 NX FPGAs“. [intel.de/content/www/de/de/products/programmable/fpga/stratix-10/nx.html](https://intel.de/content/www/de/de/products/programmable/fpga/stratix-10/nx.html).
- <sup>16</sup> Microsoft. „Was sind Field Programmable Gate Arrays (FPGA) und wie werden sie bereitgestellt?“, März 2020. <https://docs.microsoft.com/de-de/azure/machine-learning/how-to-deploy-fpga-web-service>. Nutzung mit Genehmigung von Microsoft.
- <sup>17</sup> Siehe beispielsweise: Intel. „Simplifying Cloud to Edge AI Deployments with the Intel® Distribution of OpenVINO™ Toolkit, Microsoft Azure, and ONNX Runtime“. [intel.com/content/www/us/en/artificial-intelligence/posts/microsoft-azure-openvino-toolkit.html](https://intel.com/content/www/us/en/artificial-intelligence/posts/microsoft-azure-openvino-toolkit.html).
- <sup>18</sup> Bei SAP HANA\* Lösungen senken diese neuen Angebote die Gesamtbetriebskosten (TCO), vereinfachen die komplexen Architekturen für HA/DR und Multi-Tier-Daten und bieten 22 Mal schnellere Reload-Zeiten\*. Quelle: Microsoft. „Next Generation SAP HANA Large Instances with Intel® Optane™ Drive Lower TCO“, April 2020. <https://azure.microsoft.com/en-us/blog/next-generation-sap-hana-large-instances-with-intel-optane-drive-lower-tco/>. Nutzung mit Genehmigung von Microsoft.
- <sup>19</sup> Die Neustartzeit wurde von 50 auf 4 Minuten verkürzt. Basierend auf Tests vom 30. Mai 2018. Simulierter SAP HANA\* Workload für Version 2 des Standard Application Benchmark für SAP BW Edition für SAP HANA\* vom 30. Mai 2018. Referenzkonfiguration mit herkömmlichem DRAM: Server Lenovo ThinkSystem\* SR950 mit 8 x Intel® Xeon® Platinum 8176M Prozessor (28 Kerne, 165 W, 2,1 GHz). Der Speicher besteht aus 48 x TruDDR4\* RDIMMs (16 GB, 2.666 MHz) sowie 5 x ThinkSystem\* 2,5" PM1633a SSDs (3,84 TB Kapazität, SAS, 12 GB/s, Hot Swap) für den Datenspeicher von SAP HANA\*. Als Betriebssystem dient SUSE\* Linux Enterprise Server 12 SP3. Eingesetzt wird SAP HANA\* 2.0 SPS 03 mit einer 6 TB großen Datenbank. Durchschnittliche Startzeit bis alle Daten nach dem Vorladen aller Tabellen komplett geladen sind bei 10 Iterationen: 50 Minuten.
- Neue Konfiguration mit einer Kombination aus DRAM und persistentem Intel® Optane™ Speicher: Server Lenovo ThinkSystem\* SR950 mit 8 x Intel® Xeon® Platinum 8176M Prozessor (28 Kerne, 165 W, 2,1 GHz). Der Speicher besteht aus 48 x 16 GB TruDDR4\* 2.666 MHz RDIMMs sowie 48 x persistentem Intel® Optane™ Speicher (128 GB) und 5 x ThinkSystem\* 2,5" PM1633a SSDs (3,84 TB Kapazität, SAS, 12 GB/s, Hot Swap) für den Datenspeicher von SAP HANA\*. Als Betriebssystem dient SUSE\* Linux Enterprise Server 12 SP3. Eingesetzt wird SAP HANA\* 2.0 SPS 03 mit einer 6 TB großen Datenbank. Durchschnittliche Startzeit bis alle Daten nach dem Vorladen aller Tabellen komplett geladen sind bei 10 Iterationen: 4 Minuten (12,5-fache Verbesserung).
- <sup>20</sup> Intel. „Baidu Feed Stream Services Restructures Its In-Memory Database with Intel® Optane™ Technology“. <https://newsroom.intel.com/wp-content/uploads/sites/11/2019/08/baidu-feed-case-study.pdf>.
- <sup>21</sup> Baidu benötigte Caching-Laufwerke, um kontinuierlich sequenzielle Daten an Bandlaufwerkstapeln zu liefern. Sechs NAND-Laufwerke lieferten 36,75 PBW während zwei Intel® Optane™ SSDs 164 PBW lieferten. Quelle: Intel. „Improve Tape-Backup Speeds with Intel® Optane™ SSDs“, Juli 2020. [intel.de/content/www/de/de/products/docs/storage/baidu-improves-tape-backup-case-study.html](https://intel.de/content/www/de/de/products/docs/storage/baidu-improves-tape-backup-case-study.html).

In Leistungstests verwendete Software und Workloads können speziell für die Leistungseigenschaften von Intel® Mikroprozessoren optimiert worden sein.

Leistungstests wie VYSmark\* und MobileMark\* werden mit spezifischen Computersystemen, Komponenten, Softwareprogrammen, Operationen und Funktionen durchgeführt. Jede Veränderung bei einem dieser Faktoren kann abweichende Ergebnisse zur Folge haben. Als Unterstützung für eine umfassende Bewertung ihrer geplanten Anschaffung sollten Sie zusätzliche Informationen und Leistungstests heranziehen – auch im Hinblick auf die Leistung des betreffenden Produkts in Verbindung mit anderen Produkten. Ausführlichere Informationen finden Sie unter [www.intel.com/performance](https://www.intel.com/performance).

Die Leistungsergebnisse basieren auf Tests, die zu dem in den Konfigurationen angegebenen Datum durchgeführt wurden, und spiegeln möglicherweise nicht alle öffentlich erhältlichen Sicherheitsupdates wider. Konfigurationsdetails finden Sie im Backup. **Kein Produkt und keine Komponente bietet absolute Sicherheit.**

Unter Umständen können Intel® Compiler bei Optimierungen, die nicht für Intel® Mikroprozessoren spezifisch sind, auch bei Mikroprozessoren anderer Hersteller denselben Optimierungsgrad erzielen. Zu diesen Optimierungen gehören Befehlsätze für SSE2, SSE3 und SSE3E sowie weitere Optimierungen. Intel übernimmt keine Garantie für die Verfügbarkeit, Funktionalität oder Wirksamkeit von Optimierungen für Mikroprozessoren, die nicht von Intel hergestellt wurden. Mikroprozessorabhängige Optimierungen in diesem Produkt sind für die Anwendung in Verbindung mit Intel® Mikroprozessoren bestimmt. Bestimmte, nicht für die Intel® Mikroarchitektur spezifische Optimierungen sind Intel® Mikroprozessoren vorbehalten. Entnehmen Sie weitere Informationen zu den spezifischen Befehlsatzerweiterungen, auf die dieser Hinweis zutrifft, bitte den entsprechenden Benutzer- und Referenzhandbüchern.

Kosten und Ergebnisse können variieren.

Intel hat keinen Einfluss auf und keine Aufsicht über die Daten Dritter. Sie sollten andere Quellen heranziehen, um die Richtigkeit zu beurteilen.

Intel® Technik kann entsprechend geeignete Hardware, Software oder die Aktivierung von Diensten erfordern.

© Intel Corporation. Intel, das Intel Logo und andere Intel Markenbezeichnungen sind Marken der Intel Corporation oder ihrer Tochtergesellschaften. \*Andere Marken oder Produktnamen sind Eigentum der jeweiligen Inhaber.