

Digitale Infrastruktur - Von der Cloud bis zur Edge

Kurz & klar

- Das Cloud-Computing ist eine der prägendsten technologischen Entwicklungen der letzten Jahre. Die Migration auf die Cloud ermöglicht es Unternehmen, erhebliche Vorteile bei Zusammenarbeit, Automatisierung, Skalierung, Innovation und Agilität zu realisieren. Dieser Trend dürfte in den kommenden Jahren weiter anhalten.
- Alle Segmente des Cloud-Marktes werden mit deutlich zweitstelligen Wachstumsraten zulegen. Die Dominanz der Big3 (AWS, Azure und Google Cloud) dürfte für die europäischen Anbieter selbst auf ihrem Heimatmarkt schwer zu brechen sein.
- Die globale Datenexplosion in Kombination mit erhöhten Latenzanforderungen von Anwendungen der neuen Generation erfordert einen neuen Ansatz der Datenverarbeitung. Der neue Trend „Edge Computing“ löst die Problematik der Verarbeitung der riesigen Datenmengen und des Datentransfers.

Cloud Computing – Die Erfolgsstory geht weiter

- Cloud Computing setzt neues Paradigma: technische Beschränkungen nicht mehr entscheidend

Kaum ein anderes Thema in der Technologie-Welt wurde in den vergangenen Jahren so intensiv diskutiert wie das Cloud Computing. Der Begriff „Cloud Computing“ geht auf Prof. Chellappa von der Goizueta Business School zurück, der ihn erstmals 1997 benutzte. Dabei beschrieb er Cloud Computing als neues Paradigma, bei dem die IT-Grenzen durch ökonomische Kalküle und nicht mehr durch technische Beschränkungen bestimmt werden. Eine heute weit verbreitete Definition liefert das US-amerikanische National Institute of Standards and Technology (NIST). Danach ist Cloud Computing der omnipräsente und komfortable Zugriff auf einen Pool von Ressourcen über ein Netzwerk, der von mehreren Anwendern genutzt werden kann. Zu den bereitgestellten Ressourcen zählen Netzwerke, Speicherplatz, Rechenleistung, Anwendungen und weitere Dienste, welche unverzüglich ohne menschliche Interaktion zwischen Anwender und Provider an den tatsächlichen Bedarf angepasst genutzt werden können. Nutzer erwarten dabei die ständige Verfügbarkeit dieser gemeinsam genutzten (theoretisch unendlichen) Ressourcen.

- Die Cloud erlaubt ein hohes Maß an Skalierbarkeit

Die Vorteile dieser Technologie für Unternehmen sind vielfältig. Neben der zeitnahen Skalierbarkeit von IT-Leistungen an den Bedarf zählen auch eine erhöhte organisatorische Flexibilität sowie ein reduzierter Administrationsaufwand (Betrieb- und Wartung von IT-Ressourcen entfallen) zu den Pluspunkten. Ebenfalls entfallen Investitionskosten für Server-Hardware. Aus fixen Investitionskosten und -risiken werden variable Kosten, wobei Kosten und Nutzen dem tatsächlichen Bedarf entsprechen. Investitionen in Überkapazitäten für Lastspitzen müssen entsprechend nicht vorgenommen werden. Zudem ermöglicht auch der zeit-, ort- und geräteunabhängige Zugriff neue Arbeitsmethoden.

Bei den Cloud-Varianten unterscheidet man verschiedene Arten der Bereitstellung. Dabei findet die Kategorisierung danach statt, von wem die Dienstleistungen zur Verfügung gestellt werden bzw. an welche Zielgruppe sie sich richten. Dabei haben sich über die letzte Dekade die drei Hauptkategorien Public Cloud, Private Cloud und Hybrid Cloud (bzw. Multi-Cloud) herausgebildet.

- ▶ Die Public Cloud stellt die höchste Form der Standardisierung dar

Die **Public Cloud** ist ein Angebot eines frei zugänglichen Anbieters, der seine Dienste offen über das Internet zugänglich macht. Klassische Beispiele sind das Speichern von Daten in der Cloud oder die kostenfreie Nutzung von E-Mail-Diensten. Genauso stellen Anbieter auch Rechenkapazität oder Speicherplatz für Unternehmensdaten, aber auch Software wie ERP- und CRM-Systeme, sowie Office-Anwendungen über das Internet bereit, auf die Unternehmen vollumfänglich zugreifen können. Darüber hinaus können Unternehmen mit der Public Cloud deutlich flexibler arbeiten, da die gebuchten Services sich nach Bedarf und mit wenigen Klicks aufstocken oder reduzieren lassen. Falls man z.B. zum einen Stichtag einen Rechnungsabschluss oder Reporting erstellen muss, kann man mehr Kapazität für diesen Zeitraum einfach hinzu buchen. Entsprechend ist die Public Cloud die günstigste Variante beim Cloud-Computing. Denn der Provider verarbeitet die Services von allen Kunden auf derselben standardisierten Infrastruktur, die er besser als bisher auslastet. Mehrere Nutzer teilen sich also die gleichen Ressourcen. Durch diese Skaleneffekte kann der Cloud-Provider Anwendungen besonders günstig bereitstellen.

- ▶ Die Private Cloud bietet Kunden mehr Exklusivität und bessere Sicherheitsstandards

Die **Private Cloud** kommt zum Einsatz, wenn Unternehmen aus Gründen von Datenschutz und Sicherheit ihre Dienste weiterhin selbst betreiben wollen und ausschließlich ihren eigenen Mitarbeitern zugänglich machen wollen. Dafür stellt der Anbieter dezidierte Ressourcen bereit. Der Zugriff erfolgt über das Internet oder über ein privates internes Netzwerk. Zudem sind die Dienste stärker auf die Belange des Unternehmens zugeschnitten und es lassen sich auch exklusivere Leistungen bereitstellen. Mit diesen lassen sich auch schnell neue Geschäftsfunktionen entwickeln und in Betrieb nehmen. Man „hostet“ die Infrastruktur der Private Cloud direkt vor Ort im Unternehmen oder im Rechenzentrum eines Service-Anbieters, was ein höheres Maß an Sicherheit gewährleistet. Die hohen Sicherheitsstandards und die Exklusivität bedeuten für den Dienstleister natürlich mehr Aufwand und das schlägt sich in einem höheren Preis im Vergleich zur Public Cloud nieder.

- ▶ Hybrid-Cloud - das Beste aus beiden Welten?

Die **Hybrid Cloud** stellt eine Mischform aus einer oder mehreren Public und Private-Clouds dar. Das Unternehmen entscheidet, welche Software und welche Daten in der Private Cloud und welche in der Public Cloud angesiedelt werden. So laufen bestimmte Services bei öffentlichen Anbietern über das Internet, während datenschutzkritische Anwendungen oder die Daten von gesetzlich regulierten Unternehmen auf einer Private Cloud bleiben. Der Nutzer verbindet die Kosteneffekte aus der Public Cloud mit den Sicherheitsansprüchen von Private Clouds.

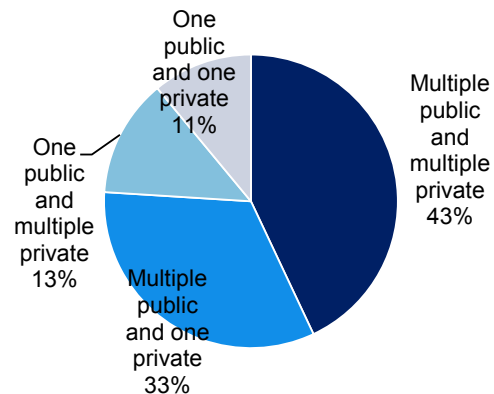
Bei einer **Multi-Cloud** liegt immer mehr als eine Cloud-Implementierung des gleichen Typs (Public oder Private) von verschiedenen Anbietern vor. Eine Multi-Cloud kann sich also aus mindestens zwei Public-Clouds oder zwei Private Clouds zusammensetzen. Dagegen besteht eine Hybrid-Cloud immer aus einer Public-Cloud und einer Private-Cloud gleichzeitig, die eine einzelne Einheit darstellt. Eine Multi-Cloud hingegen ist keine einzelne Einheit, sondern eine Serie von unabhängigen Einheiten, die unter eine zentrale Verwaltung gestellt werden. Daten und Prozesse neigen dazu, sich in einer hybriden Umgebung zu vermischen, während in einer Multi-Cloud die Nutzung typischerweise im Silo der eigenen Cloud bleibt. Unternehmen nutzen die Multi-Cloud-Strategie häufig, wenn die Anforderungen einen einzigen Provider überfordern oder wenn Ressourcen in verschiedenen Regionen aufgrund von rechtlichen Anforderungen unterschiedlich gehandhabt werden müssen.

Ebenfalls fordern Business-Continuity-Pläne Unternehmen oftmals dazu auf, Anwendungen auf verschiedenen Clouds zu platzieren.

- Die Hybrid-Cloud ist der favorisierte Ansatz für Unternehmen

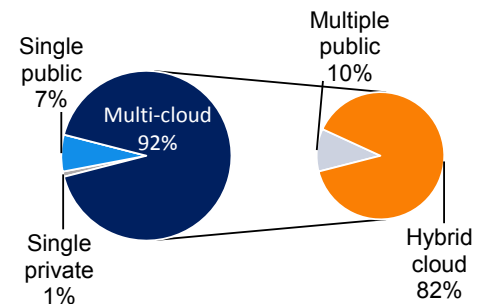
Laut dem Flexera Cloud-Report nutzen derzeit 92% der Unternehmen einen Multi-Cloud-Ansatz, wovon wiederum 82% eine Hybrid-Cloud-Strategie verfolgen. Durchschnittlich sind in Unternehmen drei Public Clouds und drei Private Clouds im Einsatz. Außerdem nutzen 76% der Unternehmen mindestens zwei Cloud-Anbieter.

Oftmals ist mehr als ein Cloud-Typ im Einsatz
Angaben in Prozent



Quelle: Flexera Cloud Report 2021, BayernLB Research

Unternehmen setzen primär auf die Hybrid-Cloud
Angaben in Prozent



Quelle: Flexera Cloud Report, BayernLB Research

- Infrastruktur, Plattformen und Software sind die drei Hauptkategorien von Services

Eine Cloud ohne entsprechende Services ist heutzutage undenkbar. Bei den Diensten handelt es sich um Infrastrukturen, Plattformen oder Software, die Nutzern über das Internet zur Verfügung gestellt werden. Dabei werden drei Haupttypen von „As-a-Service“-Lösungen unterschieden: Infrastructure as a Service (IaaS), Platform as a Service (PaaS) und Software as a Service (SaaS). Alle drei ermöglichen den Fluss von Benutzerdaten über das Internet zu den Systemen des Anbieters und zurück. Sie unterscheiden sich jedoch durch die Art der Bereitstellung und Verantwortlichkeit.

IaaS (Infrastructure as a Service): Die Bereitstellung von (virtualisierter) Computer-Hardware ist die Basisdienstleistung des Cloud-Computing. Rechenleistung, Speicher oder Netzwerk, gekoppelt mit Verfügbarkeit, kann individuell über eine Schnittstelle oder ein Dashboard gebucht werden und ersetzt die vor Ort vorgehaltene Hardware. Da Hardware einen erheblichen Kostenfaktor darstellt, bietet IaaS ein signifikantes Einsparpotenzial.

PaaS (Platform as a Service): Beim PaaS wird eine vollständige Entwicklungs- und Bereitstellungsumgebung in der Cloud zur Verfügung gestellt, über die verschiedenste Applikationen bereitgestellt werden können. Diese reichen von einfachen cloud-basierten Apps bis hin zu ausgereiften cloud-fähigen Unternehmensanwendungen. Cloud-Features wie Skalierbarkeit, Verfügbarkeit, Testumgebungen und Multiplattform-Ansätze reduzieren den Entwicklungsaufwand erheblich. Der Rollout von Updates gelingt über die Nutzung von PaaS-Diensten ebenfalls deutlich schneller.

SaaS (Software as a Service): Software und Applikationen werden als Dienstleistung zur Verfügung gestellt. Der Zugriff erfolgt zumeist über ein Web-Interface. Auf der Nutzerseite entfallen die Lizenzgebühren für die eigene Installation und der Aufwand für die Wartung. Ein weiterer wichtiger Vorteil ist die Nutzung der Services stationär und/oder mobil von unterschiedlichen Geräten aus. Die Software ist nicht mehr zwingend an einen Arbeitsplatz gebunden. Mit einer Neuanmeldung im Webtool können auch einfach neue Mitarbeiter ohne neue Lizenzen integriert werden. Zu den beliebtesten SaaS-Programmen gehören

Projektmanagement-Apps, Content-Management-Systeme (CMS), Programme für Finanzbuchhaltung, Aktenführung, eCommerce, Customer Relationship Management (CRM), Archivverwaltung und Personalplanung. Im Vergleich zu IaaS und PaaS hat der Nutzer kaum Freiheitsgrade bei Anpassungen und der Provider kümmert sich von der Bereitstellung, Datenverwaltung und Aktualisierung um nahezu alles.

Eigenbetrieb	IaaS	PaaS	SaaS
Applikation	Applikation	Applikation	Applikation
Data	Data	Data	Data
Runtime	Runtime	Runtime	Runtime
Middleware	Middleware	Middleware	Middleware
O/S	O/S	O/S	O/S
Virtualization	Virtualization	Virtualization	Virtualization
Servers	Servers	Servers	Servers
Storage	Storage	Storage	Storage
Networking	Networking	Networking	Networking

■ Nutzer verantwortlich ■ Service Provider verantwortlich

Quelle: BayernLB Research

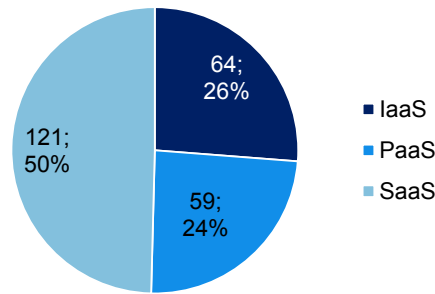
- „Pay-per-use“ und Subscriptions sind die bevorzugten Abrechnungsmodelle der Anbieter

Cloud-Provider nutzen zwei unterschiedliche Ansätze für ihr Geschäftsmodell: die nutzungsbasierte Abrechnung („Pay-per-use“) und die zeitbasierte Abrechnung (Subscriptions). Bei ersterem Modell zahlt der Nutzer eines Dienstes für das Maß der Nutzung wie z.B. für die genutzte Rechenleistung oder in Anspruch genommene Bandbreite. Subscription-Modelle basieren häufig auf monatlichen oder jährlichen Zeiträumen und stellen die Leistung unabhängig von der eigentlichen Nutzung für den Zeitraum der Zahlung bereit. Beide Modelle können auch gemeinsam genutzt werden, etwa eine Subscription als Basisleistung und optional hinzubuchbare Leistung für Leistungsspitzen. Zum Beispiel werden Maschinendaten dauerhaft in der Cloud gespeichert, während die Auswertung aber als „Pay-per-use“ nur einmal die Woche erfolgt.

- SaaS stellt das größte Segment dar

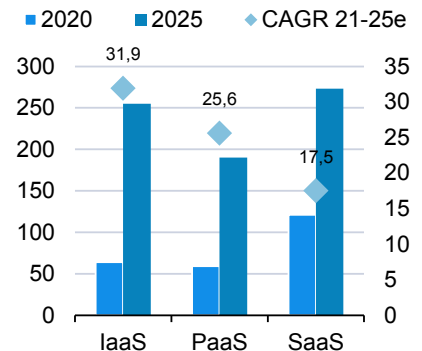
Durch die zunehmende Verwendung des Application Programming Interface (API), die als eine Programmierschnittstelle dazu dient, Informationen zwischen einer Anwendung und einzelnen Programmteilen standardisiert auszutauschen, konnten sich cloud-basierte SaaS-Lösungen in die IT-Infrastruktur des Nutzers problemlos einbinden lassen und damit die Nachfrage in diesem Segment stark steigen lassen. Entsprechend der höheren Reife des SaaS-Segments wird in den nächsten Jahren zwar mit einem geringeren Wachstum als bei IaaS oder PaaS gerechnet, jedoch vergrößert sich laut Gartner das Volumen von 121 auf 270 Mrd. USD im Jahr 2025 substantiell. Die geringeren Anfangsinvestitionen für Kunden, niedrigere langfristige Wartungskosten und erhöhte Flexibilität sollten den Cloud-Einsatz und Investitionen in Cloud-Lösungen im Allgemeinen weiter vorantreiben, was sich letzten Endes auch in weiterhin deutlich zweitstelligen Wachstumsraten in den Segmenten widerspiegeln sollte.

SaaS ist mit Abstand das größte Segment
Ausgaben in Mrd. USD für 2020



Quelle: Gartner, BayernLB Research

Wachstum in allen Segmenten bleibt hoch
Ausgaben in Mrd. USD (links), CAGR in % (rechts)

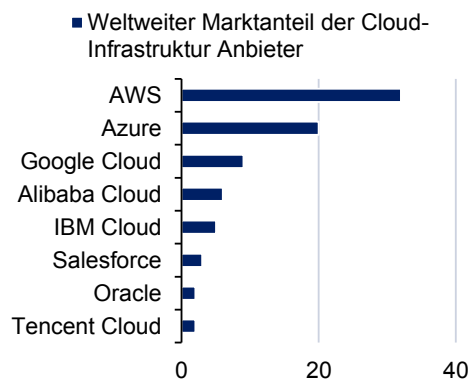


Quelle: Gartner, BayernLB Research

- Die Big3 aus den USA kontrollieren den globalen Cloud-Markt

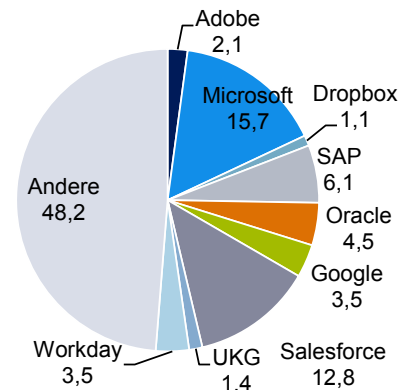
Unter den Cloud-Providern sind die US-amerikanischen Big3 weltweit führend. So hat Amazon Web Services (AWS) laut Synergy Research einen Marktanteil von etwa 33% am weltweiten IaaS/PaaS-Markt vor Microsoft Azure mit ca. 20% und Google Cloud mit ca. 10%. Der weltweite Marktanteil von europäischen Anbietern wie Deutsche Telekom oder OVHcloud liegt gerade einmal bei 1%. Auf dem SaaS-Markt sind SAP und Sage die einzigen Nicht-US-Unternehmen unter den 20 weltweit führenden Anbietern, die im Jahr 2020 mit einem Marktanteil von 6,1% bzw. 0,6% auf den Plätzen 3 bzw. 20 rangieren. Microsoft und Salesforce, die Nr. 1 bzw. Nr. 2, hatten dagegen Marktanteile von 15,7% bzw. 12,8%.

Die Dominanz der US-Provider ist sehr deutlich
Marktanteil IaaS/PaaS + hosted private Cloud in Prozent



Quelle: Synergy, BayernLB Research

Nur SAP kann den US-Anbietern Paroli bieten
Weltweiter Marktanteil SaaS in Prozent



- US-Hyperscaler dominieren auch den europäischen Markt

Jedoch nicht nur auf globaler Ebene sind die US-Hyperscaler (die Bezeichnung stammt aus der Fähigkeit, IT-Ressourcen in hohem Maß skalieren zu können) dominierend. Laut Synergy weisen die drei großen Cloud-Anbieter in Europa zusammen einen Marktanteil von 69% auf. Unter den europäischen Anbietern ist die Deutsche Telekom mit einem Anteil von 2% am europäischen Markt führend, gefolgt von OVHcloud, SAP und Orange. Die europäischen Cloud-Anbieter versuchen auf dem Markt Fuß zu fassen, indem sie sich auf Kundensegmente und Anwendungsfälle konzentrieren, die strengere Anforderungen an die Datensouveränität und den Datenschutz stellen. Die europäische Cloud-Initiative Gaia-X versucht diesen Ansatz zu bündeln. Das Ziel des Projekts ist der Aufbau einer Dateninfra-

struktur aus Europa, die die Aspekte der Datensicherheit, Datensouveränität und Portabilität der Daten (über standardisierte Schnittstellen) hervorhebt. In der Idealvorstellung sollen sich Unternehmen die Dienste nach Kriterien wie Firmensitz, Standort des Rechenzentrums oder Erfüllung regulatorischer Rahmenbedingungen (z.B. DSGVO) über einen globalen Katalog auswählen und diesen auch relativ einfach wechseln können. Damit soll auch der sog. Vendor Lock-in-Effekt aufgeweicht werden. Dieser beschreibt das Phänomen, dass Nutzer dauerhaft bei nur einem Anbieter verbleiben, wenn sich die Wechselhürden als zu hoch erweisen.

Die Dominanz der US-Hyperscaler ist in jedem europäischen Land deutlich sichtbar
Ranking nach Marktanteilen

Rang	Europa Gesamt	Großbritannien	Deutschland	Frankreich	Niederlande	Rest von Europa
Leader	AWS	AWS	AWS	AWS	AWS	AWS
#2	Azure	Azure	Azure	Azure	Azure	Azure
#3	Google	Google	Google	OVH	Google	Google
#4	IBM	IBM	Deutsche Telekom	Orange	KPN	IBM
#5	Salesforce	Rackspace	IBM	Google	IBM	Salesforce
#6	Deutsche Telekom	Salesforce	Oracle	IBM	Oracle	Swisscom

Quelle: Synergy, BayernLB Research

- Die Ausgaben für die Cloud-Technologien haben noch lange nicht das Niveau der traditionellen IT-Ausgaben erreicht

Das stark steigende Datenvolumen auf Websites und in mobilen Apps, der steigende Fokus auf die Bereitstellung kundenorientierter Anwendungen zur Steigerung der Kundenzufriedenheit und die wachsende Notwendigkeit, Investitionsausgaben und Betriebsausgaben zu kontrollieren und reduzieren, sind einige der Faktoren, die das Wachstum der neuen Technologien vorantreiben. Aufstrebende Technologien wie Big Data und Künstliche Intelligenz gewinnen an Zugkraft und führen zu einem weltweiten Wachstum des Cloud-Computing-Marktes. Das Momentum zeigt sich auch bei den Ausgaben der letzten Dekade. Während laut Synergy Research das durchschnittliche jährliche Ausgabenwachstum für traditionelle Rechenzentren im Zeitraum 2010-20 lediglich 2% betrug, lag es im gleichen Zeitraum bei den Cloud-Diensten (IaaS, PaaS und Hosted Private Cloud) im Durchschnitt bei 52% pro Jahr. Die Befragungsergebnisse des Flexera Reports 2021 zeigen auch, dass das Potenzial der Cloud noch lange nicht ausgeschöpft ist. Sowohl bei den Workloads als auch bei den Datenbeständen werden diese zu 50% über die Cloud abgewickelt oder gelagert. Dabei geben die befragten Unternehmen weiter an, diese Raten um durchschnittlich 7 bis 8 Prozentpunkte in den nächsten 12-24 Monaten zu steigern. Schaut man sich alleine das Verhältnis der Ausgaben von traditioneller Rechnerkapazität (derzeit weltweit 625 Mrd. USD) zu Cloud-Kapazität (ca. 350 Mrd. USD) an, so sieht man, dass der Verdrängungseffekt gemessen an den Ausgaben mittelfristig weiterlaufen dürfte.

Die Ausgaben für die traditionellen Systeme liegen immer noch deutlich über denen der Cloud in Mrd. USD

Traditional IT spending		Cloud IT spending	
Data-centre systems	220	Software-as-a-Service (SaaS)	121
Infrastructure software (on-premises)	229	Hosted and Private Infrastructure services	59
Application software (on-premises)	120	Infrastructure-as-a-Service (IaaS)	64
Managed services for traditional data-centre environment	30	Platform-as-a-Service (PaaS)	59
Colocation	26	Managed services for cloud and edge environments	26
		Cloud management and security services	20
		Desktop-as-a-Service (DaaS)	1
Total	625		350

Quelle: Gartner, BayernLB Research

Edge Computing – der nächste Megatrend steht in den Startlöchern

- ▶ Edge Computing bietet einen neuen dezentralen Ansatz, der das Problem der globalen Datenexplosion bewältigen kann

In den letzten Jahren haben einige Unternehmen ihren IT-Betrieb durch die zentrale Speicherung und Verarbeitung von Daten in der Cloud konsolidiert. Die Erfordernisse durch neue Anwendungsfälle, die durch Milliarden von verteilt eingesetzten Geräten möglich wurden, haben jedoch dazu geführt, dass dieses Modell nicht mehr problemlos anwendbar ist. Durch die Datenexplosion und Komplexität der in angeschlossenen Geräten (Sensoren, Laptops, Smartphones, Maschinen, Drohnen, Wearables usw.) generierten Daten können die bisher verfügbaren Netzwerke und Infrastrukturen nicht mehr mithalten. So erzeugt ein autonom fahrendes Auto pro Tag ein Datenvolumen von rund 4 Terabyte. Ein noch größeres Volumen erzeugen z.B. Flugzeuge. So sind in den Tragflächen eines Airbus A350 rund 6000 Sensoren verbaut, die pro Tag rund 2,5 Terabyte an Daten liefern. Alleine die Sensorik in den Triebwerken misst 16 Mal pro Sekunde rund 3000 Parameter. So summieren sich die einzelnen Datenelemente auf rund 150 Mio. während eines einzigen Fluges, die analysiert werden wollen. Wenn alle von den Geräten generierten Daten an ein zentrales Rechenzentrum gesendet würden, würde dies zu Bandbreitenproblemen führen. Edge Computing bietet hier eine effizientere Vorgehensweise. Der Ansatz besteht darin, den größten Teil der Aufgaben nicht in einem zentralen Rechenzentrum zu bewältigen, sondern lokal auf Geräten (Edge Devices) zu erledigen, an denen die Daten entstehen, sofern der Anwendungsfall dies erlaubt und die informationstechnischen Anforderungen erfüllt werden können. Edge Computing erfolgt „am Rand“ des Netzwerkes, um die Entstehung von Engpässen bei der Datenübermittlung zu verhindern.

- ▶ Vielen Anwendungen des IoT-Ecosystems benötigen niedrige Latenzzeiten

Ein weiterer Grund für die zunehmende Implementierung der Edge-Technologie ist der Bedarf an extrem niedrigen Latenzzeiten für viele bestehende und neuartige Anwendungen. Durch die Verlagerung von Verarbeitungsressourcen näher an die Knotenpunkte können Edge-Lösungen eine geringere Reaktionszeit (Latenz) von weniger als 20 Millisekunden (ms) bieten (im Vergleich zu 50-150 ms bei herkömmlichen Strukturen). Die Latenzzeit ist für datenlastige Anwendungen und Dienste der nächsten Generation von sehr hoher Bedeutung. So werden aktuell erzielbare Latenzzeiten für neue Technologien wie z.B. Autonomes Fahren, IoT, Industrieautomatisierung/Robotik, Augmented- und Virtual Reality (AR/VR), Echtzeit-Analytik, Smart Cities und Cloud-Gaming kaum ausreichen. Die VR-Technologie, der zuletzt vor dem Hintergrund des Metaverse-Hype mehr Beachtung zugekommen ist, zeigt die Bedeutung von Edge Computing sehr gut. Laut Qualcomm garantiert erst eine Latenz von unter 15 ms eine problemlose Benutzung von VR-Anwendungen. Wenn die Latenz höher ist, kann das menschliche Gehirn die Verzögerung nicht richtig verarbeiten. Dies führt zu der sog. motion sickness (diese wird manchmal auch als VR-Krankheit bezeichnet), die bei vielen Nutzern von VR-Brillen zu einer Übelkeit führt.

- ▶ Edge Computing wird für zeitkritische Anwendung unerlässlich sein

Bestimmte künftige Anwendungen werden sogar eine Latenzzeit von unter 1 ms erfordern. Darunter fallen die hochsensible Bewegungssteuerung in der Industrieautomatisierung, die V2X-Kommunikation (Vehicle To Everything Communication) sowie die sog. Ultra-Reliable and Low-Latency Communications (URLLC stellt ein Anwendungsprofil des 5G-Mobilfunknetzes dar), die z.B. für den Einsatz von automatisierten Prozessen der Industrie 4.0 oder Echtzeit-Remote-Verfahren der Medizintechnik verwendet werden. Edge Computing und eine verteilte Cloud-Architektur werden notwendig sein, um diese Latenzanforderungen zu erfüllen.

Übersicht von diversen Anwendungen nach Latenzzeiten

< 100ms	< 50ms	< 10ms	< 1ms
<ul style="list-style-type: none"> Fahrzeugsicherheits-Apps (Kontextbewusstsein für Warnungen) Assistiertes Fahren mit menschlichem Fahrer als Kontrolle Flugbewegungssteuerungssystem (<60ms) und Flügelüberwachungs- und -steuerungssystem (<100ms) 4G LTE Datenrate 	<ul style="list-style-type: none"> Cloud-Gaming und professioneller eSport für Fortgeschrittene (20ms) Kognitive Unterstützung (20-40ms) Virtuelle Realität (10-20ms) Kooperatives Fahren (20ms) Roboterfernsteuerung mit haptischem Feedback (25ms) 	<ul style="list-style-type: none"> Gemeinsame haptische virtuelle Umgebung Telemedizinische Anwendungen (Ferndiagnose und Rehabilitation) Erweiterte Realität Fernunterricht und -schulung mit haptischem Feedback (medizinische Fernschulung) Intelligentes Stromnetz (3ms) Hochpräzise Prozessautomatisierung (5ms) 	<ul style="list-style-type: none"> Echtzeit-Fernsteuerung & Telepräsenz mit synchroner haptischer Rückmeldung Bewegliche Industrieroboter Industrielle Regelkreissysteme (1ms-Zyklen) Autonomes kooperatives Fahren (V2V-, V2X-Kommunikation) Intelligente Stromnetze: synchrone Kopplung von Stromversorgungen (<1ms)

Quelle: IEEE Computer Society, BayernLB Research

- Eine geringere Latenz bedeutet für viele Geschäftsmodelle auch mehr Umsatz

Eine geringe Latenzzeit gibt auch Unternehmen die Möglichkeit Kunden zu gewinnen oder zu halten. Laut einer IDC-Umfrage sind 75% der Unternehmen der Meinung, dass eine nicht-zeitnahe Reaktion auf Daten ihre Geschäftsmöglichkeiten beeinträchtigt. Vor zehn Jahren behauptete Amazon, dass jede 100 ms Latenz 1% des Umsatzes koste. Google stellte fest, dass eine zusätzliche Latenzzeit von 0,5 Sekunden auf der Suchseite zu einem Rückgang des Datenverkehrs um 20% führe. Eine Studie von Akamai aus dem Jahr 2017 zeigte, dass 100 ms Latenz bei der Website-Ladezeit die sog. conversion rates (Prozentsatz der Website-Besucher, die eine gewünschte Aktion auf der Website durchführen wie z.B. Werbung durchklicken oder Artikel in den Warenkorb legen) um 7% belasten. Eine Verzögerung von zwei Sekunden bei der Ladezeit einer Website erhöht die Absprungrate (der Prozentsatz der Besucher, die die Website betreten, sie aber wieder verlassen, ohne jemals zu einer 2. Seite gelangt zu sein) um 103%. 53% der Nutzer von mobilen Geräten verlassen eine Seite, wenn sie länger als drei Sekunden zum Laden braucht.

Beispiel eines Use Case für den Einsatz von Edge Computing im Einzelhandel

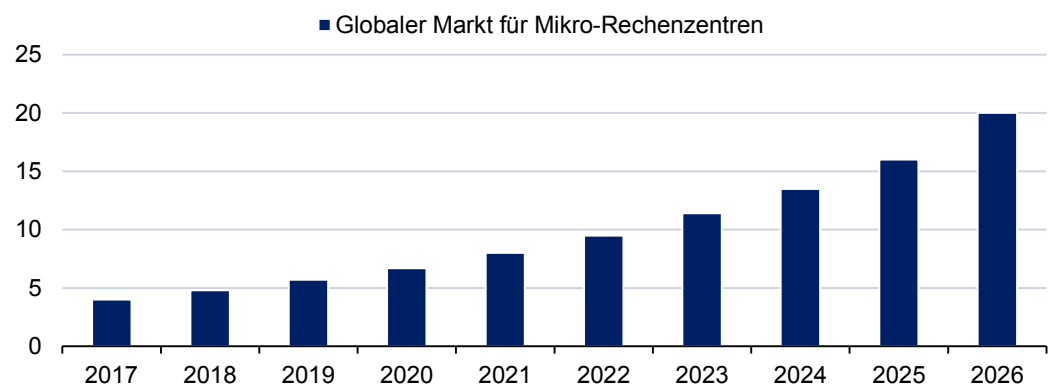
Anwendungen für den Einzelhandel erzeugen riesige Datenmengen aus Kassensystemen, Warenbeständen, Sicherheitsvideos und anderen Geschäftsaktivitäten. Edge Computing kann dabei helfen, Daten zu analysieren, Probleme zu identifizieren sowie längerfristige Verkaufstrends und Geschäftsmöglichkeiten zu erkennen. Dafür werden teils bestehende Systeme mit Edge-Rechenzentren unterstützt, sodass eine erste Analyse direkt am Ort der Datenerfassung erfolgen kann. Edge-Lösungen helfen z.B. die Verfügbarkeit der Produkte am Point-of-Sale (PoS) zu verbessern, da Abverkaufdaten in Echtzeit analysiert werden und somit die Lieferkette mit stets aktuellen Daten gefüttert wird. Die optimierte Logistik hilft auch dabei, witterungsbedingte, regionale oder saisonale Kundenpräferenzen schneller zu erkennen und somit den Verkauf zu optimieren. Mit den Edge-Systemen wird zudem eine ständige Verfügbarkeit von Kunden- und Produktdaten am PoS erreicht, sodass sich auch kurzfristige Marketingkampagnen gut umsetzen lassen. Des Weiteren können Einzelhändler auch Sensoren im Geschäft anbringen, um das Bewegungsmuster der Kunden

z.B. über Wärmekarten zu erkennen und so den Raum effizienter auszugestalten. Abgesehen von den betrieblichen Abläufen können Einzelhandelsunternehmen Edge Computing nutzen, um das Kundenerlebnis (Customer Experience) zu erhöhen. Dies beginnt bereits vor Betreten des Ladens, wenn der Kunde über sein Smartphone identifiziert wird und entsprechend seinem Kundenprofil z.B. spezielle Rabattaktionen und personalisierte Produktempfehlungen erhält. Interaktive Live-Displays, an denen der Kunde vorbei geht, zeigen ihm Neuheiten und Angebote auf Basis seiner früheren Einkäufe oder Präferenzen. Intelligente Umkleidekabinen lassen sich mit Augmented-Reality-Spiegeln ausstatten, damit Kleidungsstücke "anprobiert" werden können, ohne sie tatsächlich anzuziehen. Der Kunde kann die Farbe ändern, eine andere Größe ausprobieren und die Verfügbarkeit überprüfen. Auch beim Bezahlvorgang erhält die Edge-Technologie Einzug. Während das kontaktlose Bezahlen über RFID-Technologie schon als Standard gilt, treibt die "just walk out"-Technologie die Entwicklung noch weiter. In einem modernen Shop kann sich der Kunde innerhalb des Geschäftes mit seinem Smartphone anmelden. Der ausgewählte Warenkorb wird durch RFID-Chips, Kameras und Sensoren erfasst. Das heißt, jedes Mal, wenn ein Kunde ein Produkt aus einem Regal nimmt, wird dies unmittelbar auf seinem Smartphone vermerkt. Der Kunde verlässt den Laden ohne an einer Kasse zu bezahlen. Die Bezahlung erfolgt automatisiert per Mobile Payment auf Basis einer fälschungssicheren Blockchain-Anwendung im Edge-Rechenzentrum. Das Personal ist nur noch für die Auffüllung der Waren und die Beantwortung von Kundenfragen zuständig.

- Wachsende Anzahl von Edge-Lösungen werden den Markt für Mikro-Rechenzentren deutlich beflügeln

Im Vergleich zu einem klassischen Cloud-Rechenzentrum ist ein Edge-Rechenzentrum in der Regel näher am Endnutzergerät und hat in der Regel eine kleinere Kapazität von 50 bis 150 kW Kapazität. Diese sog. Mikro-Rechenzentren (Micro Data Center; MDC) enthalten CPU oder KI-Rechenbeschleuniger (wie FPGA, GPU, TPU), Speicher (HDD oder SSD) und das für die Ausführung von Anwendungen erforderliche Netzwerk. Ein MDC enthält meist weniger als zehn Server und weniger als 100 virtuelle Maschinen, die in einem 19-Zoll Gehäuse mit nur einem Server-Rack Platz finden. Ein größeres Rechenzentrum kann auch aus einer Reihe kleinerer Module aufgebaut werden, die insgesamt die Größe eines Containers erreichen können. Aufgrund ihrer unterschiedlichen Größe und Modularität gibt es keine universelle Form für ein Edge-Rechenzentrum, da jede Branche unterschiedliche Anpassungen für Stromversorgung, Kühlung, Sicherheits-, Überwachungs- und Brandschutzmaßnahmen vornimmt. Die Hardwaresysteme verfügen über die gleiche Grundarchitektur und die gleichen Softwarefunktionen wie in den zentralen Cloud-Rechenzentren. Der Markt für solche Mikro-Rechenzentren wird sich sehr dynamisch entwickeln. Es werden durchschnittliche Wachstumsraten von ca. 20% jährlich erwartet, bei einem weltweit zu erwartendem Marktvolumen von ca. 20 Mrd. USD im Jahr 2026.

Der Aufstieg der Edge-Technologien wird die Nachfrage nach Mikro-Rechenzentren steigern
Marktvolumen in Mrd. USD



Quelle: GMI, BayernLB Research

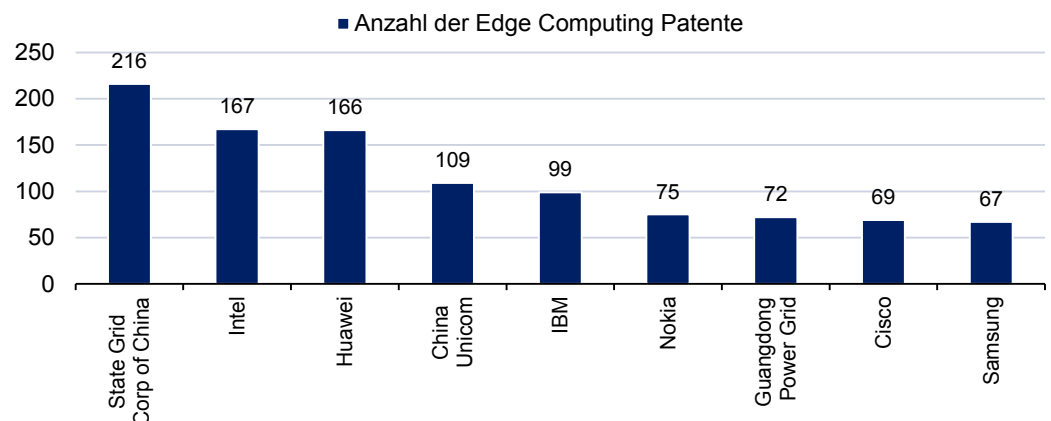
- ▶ Der Markt für Edge Computing gestaltet sich heterogener

Der Edge-Computing-Markt wird bislang nicht so eindeutig von den Hyperscalern dominiert wie der Markt für die Public Clouds. Zwar finden sich unter den Top-Playern AWS, Microsoft Azure und Google Cloud (diese bieten Cloud-IoT Edge-Lösungen an), jedoch gestaltet sich der Gesamtmarkt mit weiteren Unternehmen wie Huawei, ZTE, Dell, IBM, EdgeConneX, MobileEdgeX, Intel, Adlink Technology, Schneider Electric, Equinix, Vapor IO, Cisco, Nokia und ClearBlade deutlich heterogener. Dies liegt u.a. daran, dass klassische Cloud-Provider Konkurrenz von Telekombetreibern und Netzwerkausrüstern bekommen, die bei der Übertragungstechnologie bereits über eine hohe Kompetenz verfügen. Anbieter von drahtloser Kommunikationsinfrastruktur erweitern ihre Infrastruktur relativ einfach durch den Einsatz von Mikro-Datenzentren an Mobilfunkmasten oder Routern.

- ▶ China hat beim Know-How für Edge-Technologie eine führende Stellung

Auch bei der Technologieführerschaft zeigt sich ein unterschiedliches Bild. Während sich die Kommunikationsunternehmen Nokia und Cisco bei den Unternehmen mit den meisten Patenten für die Edge Computing-Technologie wiederfinden, haben insbesondere die chinesischen Unternehmen State Grid Corp of China und Huawei eine herausragende Stellung. Zudem ist auch China in der akademischen Forschung weit voraus. Die zehn Universitäten oder Research-Einrichtungen mit den meisten Patenten befinden sich allesamt in der Volksrepublik. Das Ecosystem in der Volksrepublik wird stark durch staatliche Förderung und Koordinations- und Kooperationsinitiativen (das Edge Computing Consortium wurde in China gegründet) im Rahmen der „Made in China 2025“-Strategie angekurbelt.

Chinesische Unternehmen sind in der Forschung der Edge-Technologien mitunter führend
Anzahl der Patente



Quelle: GreyB, BayernLB Research

„The Edge will eat the Cloud“ – Wird die Cloud, wie wir sie kennen, verschwinden?

- ▶ Cloud und Edge Computing sind gar nicht so gegensätzlich

2017 veröffentlichte Tom Bittmann, der Vizepräsident des Analyse- und Marktforschungsunternehmens Gartner, einen Blog mit dem provokanten Titel „The Edge will eat the cloud“. Seiner Argumentation nach wird die Explosion bei der Anzahl der vernetzten Geräte mit dem einhergehenden stark steigenden Datenvolumen nicht mehr mit dem klassischen Cloud-Computing zu bewältigen sein. Die Notwendigkeit von extrem niedrigen Latenzzeiten würde zusätzlich die „ältere“ Cloud-Struktur mehr und mehr obsolet machen. Auch wenn die bereits erwähnten Vorzüge in Hinblick auf Latenz und Bandbreite für den zunehmenden Einsatz des Edge Computing sprechen, erwarten wir nicht, dass die klassische Cloud in Zukunft überflüssig wird. Die beiden gegensätzlichen Ansätze zwischen Edge und Cloud (dezentral vs. zentral) sind bei näherer Betrachtung stärker verbunden als man unmittelbar vermuten mag.

- Zusammenspiel zwischen Cloud und Edge erhöht die Effizienz

Damit die Datenpakete intelligent und automatisch zwischen den MDCs und der Cloud verteilt werden können, müssen zeitkritische Informationen am Edge priorisiert werden. Dort werden die Daten zuerst gefiltert und analysiert. Unbrauchbare Daten werden aussortiert, sodass nur die relevanten Informationen in die Cloud weitergeleitet werden. Dies optimiert den Datenverkehr und belastet die Bandbreite weniger stark. Daten, bei denen keine Echtzeitverarbeitung erforderlich ist und Skalierbarkeit mehr zählt als Agilität, werden am Edge-Knoten aggregiert und an das Cloud-Rechenzentrum weitergeleitet. Während z.B. auch maschinelles Lernen und das Trainieren von Daten auf Edge-Devices stattfinden können, können längerfristige Lern-, Analyse- und Trainingsmodelle in einer zentralen Cloud ausgearbeitet und ausgeübt werden. Andersherum müssen die Algorithmen und KI-Modelle, die auf den Edge-Devices rechnen, aktualisiert und gepflegt werden. Dies ist nur möglich, wenn aktuelle Modelle von einer zentralen Infrastruktur aus gepflegt werden und diese in die Edge „gepusht“ werden. Die Cloud operiert in Zukunft dann mehr als zentraler Datenspeicher (bei dem alle Daten verschiedener Edge-Knoten zusammenlaufen) für Business Analytics, KI-Modelle und die Steuerung von Prozessen. Zusammengefasst ist es so, dass in den meisten Fällen das Cloud Computing ein effizientes „Rechnen“ am Rand des Netzwerks erst ermöglicht. Edge Computing kann die Funktionalität der Cloud definitiv erweitern und ein effizientes Zusammenspiel der Komponenten eines IoT-Ökosystems ermöglichen. Grundsätzlich stellt sie damit den nächsten Evolutionsschritt des Virtualisierungskonzepts dar.

Fazit

Das Cloud Computing bleibt auch in dieser Dekade eine Erfolgsstory. Die großen Vorteile liegen in der hohen Verfügbarkeit, der Flexibilität, der Geschwindigkeit und der inzwischen auch hochspezialisierten Services wie z.B. Business Analytics und KI-Modellierungen. Unternehmen werden ihren Daten und ihren Workload auch in Zukunft weiter in die Cloud verlagern, wobei die Hybrid Cloud (bzw. Multi-Cloud) mittelfristig die erste Wahl bei der Ressourcenverwaltung für die Unternehmen sein dürfte. Unter den Services werden alle drei Hauptkategorien (IaaS, PaaS und SaaS) in den nächsten drei Jahren mit deutlich zweistelligen Wachstumsraten zulegen. Entsprechend stark dürfen auch die Ausgaben für diese Technologie steigen, da diese mit 350 Mrd. USD auf globaler Ebene gerade einmal 56% der traditionellen IT-Ausgaben ausmachen. Die steht in Einklang mit Industrie-Umfragen. Der Forrester Predictions Report für 2021 sagte voraus, dass über 30% der Unternehmen ihre Ausgaben für die Cloud weiter erhöhen werden. Die Dominanz der US-Hyperscaler (AWS, Microsoft Azure und Google Cloud) dürfte mittelfristig nicht zu durchbrechen sein, vielmehr dürften sich europäische Cloud-Anbieter stärker auf Nischen ausrichten, die stärker den Fokus auf Datensicherheit und Datensouveränität haben.

Dem klassischen Cloud-Computing sind aber Grenzen gesetzt. Insbesondere bei Applikationen von zeitkritischer Natur wie z.B. Smart Cities, IoT, Autonomes Fahren oder VR/AR-Anwendungen werden die derzeitigen Latenzzeiten nicht ausreichen. Genau hier setzt das Konzept des Edge Computing an, welches sich in den nächsten Jahren einem starken Boom gegenübersehen dürfte. Der Einsatz von Mikro-Rechenzentren (Micro Data Center) in der Nähe der Orte wo die Daten anfallen, werden die Virtualisierung von Rechenleistung und den Einsatz von Software stärker dezentralisieren. Zudem dürften diverse Branchen wie z.B. der Einzelhandel, Rohstoffunternehmen, Logistik & Transport sowie Fertigungsunternehmen mit dem Einsatz von eingebetteten Edge-Lösungen einen spürbaren Wandel ihrer Betriebsabläufe erfahren, der teilweise bis hin zu einem neuen Kundenerlebnis führen kann. Das Eco-System im Edge Computing ist deutlich heterogener, und Unternehmen aus der Telekommunikationswelt (Provider und Netzwerkausrüster) dürften eine bedeutende

Rolle einnehmen. Die Technologieführerschaft Chinas (mit seinem größten IoT-Markt der Welt) dürfte sich aber aufgrund der oftmals großen Bedenken bzgl. Datenschutz und Industriespionage auf globaler Ebene nicht in Marktanteilen widerspiegeln. Wir glauben nicht, dass die „Edge“ das klassische Cloud Computing in Zukunft vollständig verdrängen wird. Vielmehr werden sich die beiden Ansätze ergänzen und die Allokation von IT-Ressourcen langfristig effizienter und kostengünstiger für die Anwender gestalten lassen.

miraji.othman@bayernlb.de

Allgemeiner Hinweis:

Diese Publikation ist lediglich eine unverbindliche Stellungnahme zu den Marktverhältnissen und den angesprochenen Anlageinstrumenten zum Zeitpunkt der Herausgabe der vorliegenden Information am 02.03.2022. Die vorliegende Publikation beruht unserer Auffassung nach auf als zuverlässig und genau geltenden allgemein zugänglichen Quellen, ohne dass wir jedoch eine Gewähr für die Vollständigkeit und Richtigkeit der herangezogenen Quellen übernehmen können. **Dieser Research-Bericht ist eine rein ökonomische Analyse, und kein Teil davon ist als Wertpapieranalyse oder Empfehlung zu verstehen.** Insbesondere sind die dieser Publikation zugrunde liegenden Informationen weder auf ihre Richtigkeit noch auf ihre Vollständigkeit (und Aktualität) überprüft worden. Eine Gewähr für die Richtigkeit und Vollständigkeit können wir daher nicht übernehmen. Die vorliegende Veröffentlichung dient ferner lediglich einer allgemeinen Information und ersetzt keinesfalls die persönliche anleger- und objektgerechte Beratung. Für weitere zeitnähere Informationen stehen Ihnen die jeweiligen Anlageberater zur Verfügung.

Aufgrund gesetzlicher Vorgaben (Wertpapierhandelsgesetz bzw. MiFID II) dürfen Wertpapierdienstleistungsunternehmen im Zusammenhang mit einer von ihnen erbrachten Finanzportfolioverwaltung oder unabhängigen Honorar-Anlageberatung grundsätzlich keine Zuwendungen von Dritten annehmen oder behalten. **Eine Weitergabe dieser Unterlage an Unternehmen oder Unternehmensteile, die Finanzportfolioverwaltung oder unabhängige Honorar-Anlageberatung erbringen, ist daher nur gestattet, wenn mit der BayernLB hierfür eine Vergütung vereinbart wurde.**

Die im Text genannten Finanzmarktinformationen stammen von Bloomberg und Refinitiv, soweit nicht anders vermerkt.



Miraji Othman
Senior Analyst Technology

Telefon: +49 89 2171-25888
miraji.othman@bayernlb.de

Geschäftsgebäude:
Bayerische Landesbank
Briener Straße 18
80333 München (=Paketadresse)
www.bayernlb.de