

Kapitel 1

Einleitung

Die Buchreihe *Informatik im Fokus* beschäftigt sich mit der Darstellung zeitnaher und gut verständlicher Einführungen in aktuelle Technologien. Dieses Buch will einen Überblick über Cloud Computing Architektur, Dienste und Anwendungen vermitteln, ohne Anspruch auf Vollständigkeit zu erheben. Ziel ist es, die Leserinnen und Leser auf einen einheitlichen Stand zu bringen und so eine gemeinsame Diskussionsgrundlage zu erreichen. Technische Vorkenntnisse sind dabei nicht erforderlich.

1.1 Beschreibung der Thematik

Der Begriff Cloud Computing ist zurzeit in der Informationstechnik (IT) allgegenwärtig. Was verbirgt sich dahinter? Es gibt viele Interpretationen, aber keine standardisierte oder gar einheitliche Definition. Cloud Computing erlaubt die Bereitstellung und Nutzung von IT-Infrastruktur, von Plattformen und von Anwendungen aller Art als im Web elektronisch verfügbare Dienste. Der Begriff *Cloud* soll dabei andeuten, dass die Dienste von

einem Anbieter im Internet (bzw. im Intranet eines größeren Unternehmens) erbracht werden. Die Nutzer der Cloud-Dienste können ihre eigenen Angebote wiederum selbst als Dienste im Internet bzw. Intranet anbieten.

Cloud-Ressourcen sind in der Regel virtualisiert: Der Cloud-Nutzer hat dadurch stets eine wunschgemäße, beliebige Sicht auf seine Infrastruktur und es gibt in diesem Fall keine systembedingten Abhängigkeiten oder Zwangsbedingungen für seine Anwendungen.

Cloud-Dienste sind darüber hinaus dynamisch skalierbar: Wenn eine Anwendung zusätzliche Ressourcen benötigt, können diese sofort und ohne großen Aufwand automatisch dazu geschaltet werden. Entwickler mit innovativen Ideen für neue Internet-Anwendungen müssen so beispielsweise nicht länger vorab in Hardware investieren, um ein Unternehmen zu gründen. Sie können Ressourcen flexibel von einem Anbieter beziehen und sich auf die Umsetzung ihrer Geschäftsidee konzentrieren. Wenn die Nachfrage steigt, lässt sich die Infrastruktur weitgehend automatisch an die wachsenden Anforderungen anpassen.

Cloud Computing folgt den Ideen des Utility Computing. Es wird immer die aktuell benötigte Menge an Ressourcen zur Verfügung gestellt und bezahlt. Wenn nichts genutzt wird, kostet es auch nichts. Cloud Computing hat somit eine wirtschaftliche Bedeutung, da signifikante Kostenersparnisse aufgrund der flexiblen Bereitstellung und Nutzung von Diensten möglich sind. Die vorgehaltenen abrufbaren Kapazitäten sind dabei sehr umfangreich, wodurch ein großer Skaleneffekt (*Economy of Scale*) mit einem sehr günstigen Preis/Leistungsverhältnis entsteht.

Es gibt mittlerweile zahlreiche kommerzielle Anbieter wie z. B. Amazon, Google oder Microsoft. Deren Angebote unterscheiden sich jedoch in ihrer Art: Amazon bietet virtualisierte Ressourcen zur generischen Nutzung an, während die Clouds von Google und Microsoft das Hosting von Anwendungen erlauben. Alle aktuellen Angebote sind aber zugunsten eines Wett-

bewerbsvorteils proprietär und Standards existieren nicht, womit ein schneller Wechsel des Anbieters in der Regel nicht einfach möglich ist.

Kritiker des Cloud Computing führen diese Gefahr des Vendor-Lockin neben möglichen Sicherheitsbedenken gerne ins Feld. Vor allem aus dem Kreis der etablierten IT-Manager und Abteilungsleiter wird zur Vorsicht gemahnt. Bei näherer Betrachtung scheint es hier aber in erster Linie sehr oft um die Sicherung historischer Besitzstände in den klassischen Rechenzentren zu gehen. Es sind daher zurzeit vor allem junge Unternehmen (Startups) ohne Abhängigkeiten dieser Art, die aus der neuen Technik Gewinn erzielen. Aber auch in etablierten Unternehmen, die sich dem Thema Cloud Computing stellen, bewirkt dessen Einsatz bereits nachhaltige Effizienzsteigerungen. Neben der Nutzung von *Public Clouds* etablieren sich hier unternehmensinterne *Private Clouds*.

Aus der Sicht des Kunden ergibt sich eine Reihe von Möglichkeiten, produktiver und flexibler zu arbeiten: Durch die Wahrnehmung von Angeboten im Markt ist es dem Bezieher von IT-Diensten möglich, sich gegenüber dem klassischen statischen Rechenzentrumsbetrieb zu emanzipieren. Cloud Computing entwickelt dabei eine *Kraft der schöpferischen Zerstörung* und Altes wird durch Neues ersetzt [25, 8].

Das Corporate Executive Board (CEB) in Washington hat eine Studie vorgestellt, die auf einer Befragung von 200 Business- und IT-Führungskräften basiert [62]. Darin wird prognostiziert, dass in den nächsten Jahren die klassischen IT-Abteilungen große Teile ihrer Verantwortung an das Business-Management verlieren werden und als Folge dessen auf ein Viertel ihrer Personalstärke schrumpfen. Zentrale Treiber der Veränderung sind das Cloud Computing, Soziale Medien und die Zunahme der sogenannten *Knowledge Worker* (Wissens- oder Kopfarbeiter) – gepaart mit einer Unternehmens-IT, die ihre Effizienzpotenziale weitgehend ausgeschöpft hat. Mit dem Siegeszug immer stärker

spezialisierte Cloud-Dienste wird das Business-Management zunehmend in die Lage versetzt, ohne Beteiligung der internen IT-Abteilung und an ihr vorbei selbst IT-Lösungen einzukaufen. In gleicher Weise können sich die Knowledge Worker mit ihren smarten Endgeräten und sozialen Netzen Informationen zu IT-Systemen sowie die zugehörigen Services selbst besorgen.

Damit ändert sich auch die Rolle des Leiters der IT-Abteilung: Er wird entweder zum Leiter einer internen schlanken Querschnittsabteilung oder zum Einkäufer und Manager von externen Diensten. Das klassische Rechenzentrum wandelt sich in diesem Szenario zum IT-Servicezentrum.

1.2 Definition

Obwohl es keine standardisierte, einheitliche Definition für Cloud Computing gibt, sind die grundlegenden Konzepte als auch die generellen Ziele des Cloud Computing unbestritten: Cloud Computing nutzt Virtualisierung und das moderne Web, um Ressourcen verschiedenster Art als elektronisch verfügbare Dienste dynamisch bereitzustellen. Die Dienste sollen dabei von mehreren Konsumenten verlässlich und skalierbar nutzbar sein, d. h. sowohl auf Abruf als auch nach Bedarf verfügbar sein. Aus der Sicht des Cloud-Anbieters impliziert dies in der Regel eine Multi-Mandanten Architektur und ein nutzungsabhängiges Abrechnungsmodell. Wir definieren Cloud Computing demnach wie folgt:

Unter Ausnutzung virtualisierter Rechen- und Speicherressourcen und moderner Web-Technologien stellt Cloud Computing skalierbare, netzwerk-zentrierte, abstrahierte IT-Infrastrukturen, Plattformen und Anwendungen als on-demand Dienste zur Verfügung. Die Abrechnung dieser Dienste erfolgt nutzungsabhängig.

Diese Definition legt dabei nicht fest, ob die Dienste auf Basis eines verteilten Systems oder eines einzelnen leistungsstarken Servers, z. B. eines Mainframes, erbracht werden. Dies steht im Gegensatz zum Grid Computing, wo das System immer verteilt ist. In der Regel liegt auch Cloud-Diensten eine verteilte Infrastruktur zu Grunde, dessen Management jedoch ist typischerweise zentral (und proprietär) durch einen Anbieter bestimmt. Auch dies unterscheidet das Cloud Computing vom Grid Computing, wo verteilte Knoten in der Regel autonom sind. Eine ausführliche Diskussion findet sich in [30].

Entscheidend für das Cloud Computing ist zudem dessen wirtschaftliche Bedeutung. Eine konsequente Dienste-Orientierung sowie die Nutzung von Web-Standards und des Internet als integrierte Technologie- und Geschäftsplattform positionieren das Cloud Computing für Anwendungen verschiedenster Art. Dies beinhaltet insbesondere Web-Anwendungen und modulare Dienste in verteilten Geschäftsnetzwerken und Prozessketten.

Eine viel zitierte Definition von Cloud Computing wurde vom Nationalen Institut für Standards und Technologie (NIST) in den USA erstellt [105]. Diese bestimmt fünf wesentliche Eigenschaften des Cloud Computing, drei verschiedene Klassen von Diensten und vier unterschiedliche Betriebsmodelle. Als wesentliche Eigenschaften werden benannt:

- **Diensterbringung auf Anforderung:** Dienste sind auf Anforderung und selbständig von Konsumenten ohne erforderliche menschliche Interaktion mit dem Anbieter nutzbar.
- **Netzwerkbasierter Zugang:** Dienste können netzwerkbasierend in Echtzeit durch Verwendung von Standardtechnologien abgerufen werden.
- **Ressourcen Pooling:** Ressourcen sind in Pools konsolidiert und erlauben eine parallele Diensterbringung für mehrere Nutzer (Mandanten), die dem tatsächlichen Bedarf eines jeden Nutzers angepasst ist.

- **Elastizität:** Ressourcen werden schnell und in verschiedenen, auch fein granularen Quantitäten, zur Verfügung gestellt und erlauben so die Skalierung von Systemen. Dem Nutzer gegenüber entsteht die Illusion unendlich verfügbarer Ressourcen.
- **Messbare Dienstqualität:** Dienste sind quantitativ und qualitativ messbar, so dass eine nutzungsabhängige Abrechnung und Validierung der Dienstqualität gegeben ist.

Die Klassen der Dienste und die Betriebsmodelle diskutieren wir in Kapitel 3.

1.3 Gliederung

Dieses Buch beschäftigt sich zunächst mit den Grundlagen des Cloud Computing, mit Basistechnologien wie Virtualisierung und Web Services. Es schließt sich eine Diskussion der Cloud-Architektur mit ihren Service-Bausteinen an. In den darauf folgenden Kapiteln geht es um ausgewählte Cloud-Angebote und Management-Werkzeuge. Insbesondere kommen hier auch Fra-

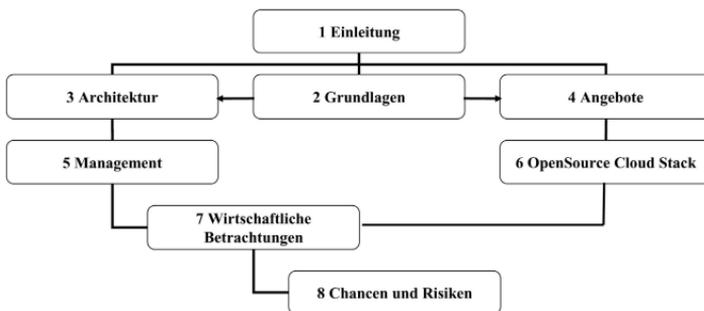


Abb. 1.1 Wegweiser zum Lesen des Buchs

gen nach Sicherheit und Datenschutz zur Sprache. Es folgt eine Betrachtung aktueller Open Source Entwicklungen wie z. B. Hadoop und Eucalyptus. Das OpenCirrusTM-Projekt von HP, Intel und Yahoo!, bei dem die Autoren als Entwicklungspartner beteiligt sind, greift diese Themen auf. Das Projekt untersucht Grundlagen der Cloud-Systeme und Cloud-Anwendungen. Es schließen sich wirtschaftliche Betrachtungen wie z. B. Kosten- und Geschäftsmodelle an. Den Abschluss bildet eine situative Bewertung des Cloud-Marktes. Im Anhang sind einige praktische Beispiele zum Umgang mit Cloud-Ressourcen bzw. Cloud-Anwendungen aufgeführt und ein Glossar fasst die Definitionen der zentralen Begriffe zusammen.

Der geneigte Leser wird das Buch mit Genuss vom Anfang bis zum Ende lesen. Die Kapitel sind aber so geschrieben, dass sie ohne Weiteres auch einzeln gelesen und verstanden werden können. Ein weniger technisch interessierter Leser kann das Kapitel 2 überspringen und sich in medias res begeben. Falls die Zeit knapp bemessen ist, gibt es einen Pfad, der sich eher an technischen Architekturfragen orientiert (Kapitel 1, 3, 5, 7, 8), und einen Pfad für eher wirtschaftswissenschaftlich interessierte Leser (Kapitel 1, 4, 6–8). Die verschiedenen Wege sind schematisch in Abbildung 1.1 aufgezeigt.