

# 1 Cloud BI & Analytics – ein Überblick

Ralf Finger · Uwe Müller

*Um die Konsequenzen und Vorteile einer Verlagerung von Business Intelligence & Analytics (BIA) in die Cloud zutreffend abzuschätzen, ist zunächst ein präzises Begriffsverständnis von Cloud Computing notwendig. Wie so oft bei neuen Technologietrends vermischen sich auch beim Thema Cloud die tatsächlichen Eigenschaften von Cloud Computing mit dem nicht gerade transparenzfördernden Marketinggetöse der Anbieter.*

*Darüber hinaus ist es für eine Nutzenbetrachtung wichtig, einen Überblick über derzeit vorhandene und entstehende Serviceangebote von Unternehmen, Kommunen und Communitys herzustellen. Hier sind die Entwicklungen hinsichtlich Open-Data-Angeboten und Data-born-in-the-Cloud von besonderer Bedeutung. Nur wenn die tatsächlichen Sourcing-Optionen und Services der Cloud sauber beschrieben und in ihren konkreten Auswirkungen klar definiert sind, lässt sich auch eine Strategie für Cloud Computing ableiten.*

## 1.1 Was ist Cloud Computing?

### Definition

Cloud Computing umschreibt IT-Infrastrukturen (u.a. Rechenkapazität, Datenspeicher, Datensicherheit, Netzkapazitäten, Entwicklungsplattformen oder auch einsatzbereite Software), die entfernt über ein Netzwerk bereitgestellt werden und in der Regel in hohem Maße Mechanismen der Virtualisierung von IT-Ressourcen nutzen. Die Bereitstellung von Cloud Computing kann als Public Cloud über öffentliche Netze und in mehrmandantenfähigen Umgebungen erfolgen. Auch ist der Aufbau von Cloud-Infrastrukturen in privaten Rechenzentren der Anwenderorganisationen möglich. Letzteres wird als Private Cloud Computing bezeichnet, wobei hier die Übergänge zu klassischen Formen des Outsourcings fließend sind. In Abgrenzung zu Cloud werden eigene Infrastrukturen beim Anwender als »On-Premises« bezeichnet.

Business Intelligence & Analytics (BIA) kann heute vollständig in der Cloud betrieben werden. Diese Definition schließt explizit alle mit diesen Systemen verbundenen Datenhaushalte mit ein.

### **Potenziale von Cloud für BI & Analytics**

Verantwortliche für BI & Analytics werden aktuell immer wieder mit zwei Anforderungen konfrontiert: Agilität und Skalierbarkeit. BIA-Anwendungen sollen sich schnell auf ändernde Bedürfnisse der Fachbereiche anpassen lassen – möglichst ohne langwierige Implementierungsprojekte. Darüber hinaus sollen BIA-Infrastrukturen im Hinblick auf Datenvolumen und Benutzeranzahlen skalieren. Allerdings müssen auch umfangreiche Leerkapazitäten vermieden werden – die Kosten sollen überschaubar und planbar bleiben.

Für all diese Anforderungen bietet Cloud-BIA heute überzeugende Lösungsansätze, sei es als eigenständige BIA-Plattform in der Cloud, als hybride Ansätze integriert in die On-Premises-Angebote der großen BIA-Plattformanbieter oder auf Basis von beispielsweise Amazon AWS, Microsoft Azure, Oracle Cloud, Google Cloud, IBM, SAP etc. Stets gilt das Leistungsversprechen: Hohe Investitionsvolumina und Großprojekte werden vermieden, die Kosten werden monatlich oder auf der Basis von Pay-per-Use-Modellen berechnet – bis hin zu minutengenaue Abrechnung von Nutzungsentgelten und Zahlung per Kreditkarte.

### **Anwendungen für Cloud BI & Analytics auf dem Vormarsch**

Nicht alle BIA-Anwendungsszenarien sind für Cloud-BIA geeignet. Pauschale Abwehrhaltungen z.B. mit Hinweis auf Datenschutz und Datensicherheit gehen jedoch fehl. Cloud-BIA muss vielmehr schon aus diesen Gründen bewusst thematisiert werden. Nur so lässt sich klar unterscheiden, in welchen Bereichen des Unternehmens Cloud-Lösungen genutzt werden dürfen und in welchen nicht.

Cloud-BIA hat sich dabei als Mittel zur Bereitstellung von Reporting- und Analysefunktionen und den damit verbundenen Inhalten über das Internet bereits etabliert. Oft als On-Demand-BIA bezeichnet, ermöglicht Cloud-BIA die standortunabhängige Bereitstellung und Verwaltung der erforderlichen BIA-Komponenten. Dies geschieht in der Regel durch Hosting-Serviceprovider.

Nach einer Anfang 2017 veröffentlichten Studie von BARC und der Eckerson Group zeigt sich, dass die Akzeptanz von Cloud-Lösungen im BIA- und Datenmanagement-Umfeld wächst [BARC & Eckerson 2017]. Die Studie zeigt ein Wachstum von weltweit um 50% zwischen 2013 und 2016 bezogen auf den Einsatz von Cloud-BIA- und Datenmanagementlösungen. Für die Anwender zählen dabei vor allem die wegfallenden Kosten für Hardware und Infrastruktur sowie der geringere Administrationsaufwand.

Mit über 50% setzen Power-User doppelt so häufig auf Cloud-Lösungen wie reine Informationskonsumenten. On-Demand-BIA-Tools bieten dabei die Mög-

lichkeit, Umgebungen zu erstellen (z. B. durch Eröffnung eines Kontos), Daten in die Umgebung zu laden und für Auswertungen und Visualisierung bereitzustellen. Am häufigsten kommen dabei Reporting- und Dashboarding-Werkzeuge zum Einsatz. Darüber hinaus finden Werkzeuge für Ad-hoc-Analysen, Datenaufbereitung, Advanced und Predictive Analytics immer mehr ihren Weg in die Cloud [BARC & Eckerson 2017].

Die bedeutenden BIA-Anbieter haben diesen Trend erkannt und bieten ihre Reporting- und Analyseplattformen in der Cloud an. In Teilen wird der volle Leistungsumfang nur noch in der Cloud angeboten bzw. die On-Premises-Lösungen werden nicht mehr weiterentwickelt. Umfangreiche Lösungen im Bereich von Cognitive Computing werden derzeit im Rahmen von Platform-as-a-Service-Angeboten (z. B. IBM Watson) zur Verfügung gestellt und können so schneller in die Analyse- und Entscheidungsprozesse des Unternehmens integriert werden.

## 1.2 Servicemodelle in der Cloud

Bei den Servicemodellen ist zwischen den klassischen Modellen der Cloud-Anbieter und darauf aufbauenden BIA-Services zu unterscheiden. BIA-Services können dabei von den Cloud-Providern selbst angeboten, von Drittanbietern als BIA-Service bereitgestellt oder von den Unternehmen selbst implementiert werden.

### 1.2.1 Typische Cloud-Service-Modelle

Bei der Gliederung von Cloud Computing werden typischerweise die folgenden Servicemodelle unterschieden:

- Infrastructure as a Service (IaaS)
- Platform as a Service (PaaS)
- Software as a Service (SaaS)

Jedes Servicemodell bietet eine unterschiedlich weitgehende Servicetiefe, mit der IT-Leistungen in der Cloud bereitgestellt werden.

Eine moderne Cloud-Strategie für BI & Analytics nutzt heute bedarfsabhängig alle drei Servicemodelle des Cloud Computing und bildet damit eine zusätzliche Sourcing-Option für IT-Leistungen.

#### **Infrastructure as a Service (IaaS)**

Bei IaaS greift der Benutzer auf IT-Basisdienste zu. Diese umfassen virtuelle Server, Storages und Netze. Der wesentliche Vorteil gegenüber traditionellem IT-Sourcing ist die Skalierbarkeit: Die Recheninstanzen können bedarfsorientiert angepasst (vertikale Skalierung) oder auch um weitere Instanzen ergänzt oder reduziert werden (horizontale Skalierung). Der Benutzer hat dabei alle Berechnungs-

gungen auf die jeweiligen virtuellen Instanzen, damit aber auch die volle Verantwortung für die Infrastruktur ab der Betriebsebene.

Für den Anwender bedeutet dies, dass On-Premises-Infrastrukturen schnell und flexibel ergänzt werden können. Dies ist z. B. für Testszenarien hilfreich, bei denen die neueste Version der Software einer On-Premises-Installation geprüft werden soll, um funktionale Unterschiede oder die Komplexität bei der Migration abzuschätzen. Mit IaaS kann dies geschehen, ohne die bestehende Infrastruktur anzutasten.

### **Platform as a Service (PaaS)**

PaaS bietet Entwicklungs- und Laufzeitumgebungen für Anwendungsentwicklung und -betrieb in der Cloud. Im Unterschied zu IaaS hat der Benutzer hier keinen direkten Zugriff auf die Rechnerinstanzen. Er betreibt auch keine virtuellen Server. In PaaS-Szenarien bringt er vielmehr seine Programm-/Ablauflogik in die Cloud. Schnittstellen in Form von Programmierschnittstellen oder komfortable GUIs dienen dabei als Unterstützung. Die Cloud-Infrastruktur regelt hierbei die erforderliche Instanziierung der Verarbeitungseinheiten und das Verteilen der zu verarbeitenden Daten sowie auf Wunsch eine automatische Skalierung an Leistungsanforderungen.

Im Kontext von Business Intelligence & Analytics finden sich solche Entwicklungsplattformen in der Cloud insbesondere bei den Datenspeichern. Hier sind relationale Datenbanken, Big Data Stores und Object Stores gut etablierte Angebote von Cloud-BIA.

### **Software as a Service (SaaS)**

SaaS stellt die Funktionalitäten einer Anwendungssoftware in der Cloud bereit. Der Endnutzer bringt hierbei weder eine Applikation in die Cloud ein, noch muss er sich um Skalierbarkeit, Datenhaltung oder IT-Ressourcen kümmern. Er nutzt lediglich die Applikation, die in der Cloud bereitgestellt wird. Anpassungen an der Applikation bewegen sich dabei in dem Rahmen einer Standardsoftware. Die IT-Administration liegt vollständig in den Händen des Servicegebers.

BIA-Umgebungen werden typischerweise durch den Einsatz von Werkzeugumgebungen aufgebaut. Diese umfassen Datenbanken, Reporting, Planungsumgebungen und Datenintegrationswerkzeuge. Die mit diesen Softwarewerkzeugen erzeugten Artefakte werden über Laufzeitumgebungen der zugehörigen Plattformen ausgeführt. Alle genannten Werkzeuge sind heute als Software as a Service in der Cloud verfügbar, sodass sich BIA-Umgebungen vollständig in der Cloud aufbauen lassen.

## Business Intelligence & Analytics as a Service

Alle typischen Komponenten einer BIA-Architektur können heute über Cloud-basierte Angebote bezogen werden. Diese sind als SaaS-Angebote ausgestaltet, wodurch der Anwender von infrastrukturbezogenen Tätigkeiten weitgehend entlastet wird. Aufgrund der Besonderheit der Cloud und der Charakteristik der Komponenten ergeben sich jedoch Eigenheiten dieser Angebote, die im Einzelnen zu bedenken sind:

- **Reporting as a Service** bezeichnet den Aufbau berichtsorientierter Anwendungen in der Cloud. Hierzu gehören formatierte Berichte, geführte Berichts-anwendungen, Dashboards und einfach zu bedienende Self-Service-Werkzeuge für die Datenanalyse. Reporting-as-a-Service-Angebote ähneln in der Funktionsweise den On-Premises-Varianten. So sind typischerweise Administrationsoberflächen zum Beispiel für die Anlage von Anwendern und die Berechtigungsverwaltung vorhanden, wie Report-Designer-Oberflächen, mit denen letztlich Anwendungen für Berichtsempfänger erstellt werden. Diese Werkzeuge müssen jedoch stets auf Daten zugreifen. Diese müssen entweder ebenfalls in die Cloud geladen werden oder werden über gesicherte Verbindungen online aus den unternehmensinternen Datenbanken gezogen. Moderne technische Konzepte erlauben mitunter, dass die Daten auch im Falle eines Onlinezugriffs das Unternehmensnetzwerk verlassen (z.B. SAP Analytics Cloud). Hierbei werden Cross-Origin-Resource-Mechanismen eingesetzt, die es Webbrowsern erlauben, auf die Ressourcen unterschiedlicher Server zuzugreifen [CORS 2018].
- Bei **Planning as a Service** werden Planungsanwendungen in der Cloud etabliert. Der besondere Vorteil von Cloud-basierten Planungslösungen ist, dass diese u.U. nur temporär benötigt werden (z.B. Planungszyklus von drei Monaten zur Grobplanung der nächsten fünf Geschäftsjahre). Außerdem können dezentrale Planer über die Cloud leicht angebunden werden, auch wenn diese nicht über das Unternehmensnetzwerk erreichbar wären. Allerdings werden Plandaten stets in einem Cloud-basierten Datenhaushalt erfasst. Des Weiteren können Planungen typischerweise nicht sinnvoll ohne zugehörige Ist-Referenzdaten durchgeführt werden. Immer werden Stammdaten (Produkte, Regionen, Kundengruppen etc.) für die Ausgestaltung der Planungsanwendung benötigt. In aller Regel sind auch historische Wertreihen als Vergleichs- oder Vorgabegrößen erforderlich. Daher ist bei Planning as a Service sorgfältig zu prüfen, ob die Vertraulichkeit der Daten eine Cloud-Lösung zulässt und falls ja, unter welchen besonderen Rahmenbedingungen dies erfolgen kann.

- **Data Warehouses as a Service (DWHaaS)** bezeichnet die Bereitstellung kompletter DWH-Services in der Cloud. So bietet z.B. Amazon Redshift ein schnelles, vollständig verwaltetes Data Warehouse, mit dem Unternehmen im Zusammenspiel mit ihren vorhandenen Business-Intelligence-Tools und mithilfe von Standard-SQL alle Unternehmensdaten analysieren können. Die Auslagerung von vollständigen Data Warehouses wird allerdings mitunter noch kritisch gesehen. Im Mittelpunkt dieser Bedenken stehen dabei oft Anforderungen an Integration, Informationssicherheit und Individualisierbarkeit. Es ist aber zu erwarten, dass neben den klassischen Cloud-Angeboten wie SaaS, IaaS und PaaS Cloud-basierte Data Warehouses in sämtlichen Branchen mehr und mehr an Bedeutung gewinnen werden. Als Gründe dafür gelten vor allem wirtschaftliche Faktoren sowie ein Zugewinn an Flexibilität und Agilität. Aber auch die typischen Bedenken gegen Data Warehouses in der Cloud dürften durch den hohen Professionalitätsgrad großer Cloud-Anbieter gepaart mit geeigneten technischen Lösungen an Bedeutung verlieren.
- Stets basieren BIA-Lösungen auf auszuwertenden Daten. Entsprechend existieren auch **Data-Integration-as-a-Service**-Angebote, die auf Daten On-Premises oder in der Cloud zugreifen und geeignete Verfahren der Datentransformation unterstützen. Für den Zugriff auf unternehmensinterne Daten werden hierfür Secure Gateways etabliert, die mit der hausinternen IT-Sicherheit abzustimmen sind. In diesem Kontext jedoch sind stets die Latenz der Datenübertragung und die Datenmenge zu prüfen. So wird in der Regel ein Online-Umzug sehr großer Datenmengen mit diesen Werkzeugen nicht im Standard durchzuführen sein. Als Alternative zum physischen Transfer von Datenträgern sind Spezialangebote ausgewählter Hersteller zu prüfen (z.B. IBM Lift).
- Mittels **Predictive as a Service** werden Werkzeuge aus Data Science, Machine Learning und Cognitive Computing über die Cloud bereitgestellt. Cloud-basierte Data-Science-Plattformen (z.B. IBM Watson Studio, SAS Viya) unterstützen dabei den kompletten Arbeitsprozess von Data-Scientisten von der Datenvorbereitung über die Modellbildung bis hin zur Bereitstellung. Fertige Prognosemodelle können auf diesen Plattformen zum Aufruf für andere Werkzeuge bereitgestellt werden. Dazu werden APIs verwendet, die von Webcalls mittels JavaScript oder auch mit aus Data Science gängigen Sprachen wie Python nutzbar sind. Einen besonderen Vorteil bietet der Funktionsreichtum Cloud-basierter analytischer Services. So können z.B. Verfahren der Bilderkennung oder des Cognitive Computing per Mausklick genutzt werden, ohne dass kostspielige Infrastrukturprojekte nötig werden. Lediglich die Frage des Datenschutzes ist wiederum aufmerksam zu prüfen.

Die obige Aufzählung zeigt die wichtigsten Funktionskreise von Cloud-BIA. Dabei ist zu beachten, dass eine Cloud-BIA-Strategie stets eine Kombination dieser Elemente sein wird. Zum Beispiel erfordern Predictive-Modelle stets Daten,

die ihrerseits über Datenintegrationsverfahren und Data Warehouses bereitgestellt werden müssen. Außerdem wird Cloud-BIA selten isoliert, sondern typischerweise mit unternehmenseigenen Komponenten eingesetzt. So können mittels logischer Data Warehouses Teile eines Data Warehouse in die Cloud verlagert werden, wodurch sich die Stärken von On-Premises und Cloud sinnvoll kombinieren lassen. Mithilfe von Datenvirtualisierungswerkzeugen können die jeweiligen Data Repositories für Auswertungen und Analysen verbunden werden. Eine solche hybride Architektur kann daher nur ausgehend von den Anforderungen des Unternehmens individuell ausgestaltet werden.

### 1.2.2 Data-born-in-the-Cloud

Digitalisierung, Internet of Things, Social Media und Cloud-basierte Geschäftsmodelle produzieren Daten, die verwertbar und überall zugänglich sein sollen. Die besondere Eigenheit vieler dieser Daten ist, dass die Datenproduzenten selbst auf der Cloud basieren. In der Konsequenz entstehen die Daten in der Cloud und sind dort unmittelbar zugreifbar. Aus diesem Grund ist die Überlegung, diese Daten auch mittels Cloud-BIA auszuwerten, naheliegend, umso mehr, wenn für die Nutzung des Service keine unternehmensinternen Daten in die Cloud bewegt werden müssen.

Diese Angebote beinhalten die Vorteile der Cloud, wie schnelle Verfügbarkeit, Skalierbarkeit und On-Demand-Zugriff, sowie die schnelle Entwicklung von Diensten basierend auf diesen Daten und Anwendungen. Wetterdaten, Social Media und Open Data sind typische Beispiele für Data-born-in-the-Cloud.

#### Wetterdaten

Das Wetter hat für Unternehmen eine hohe wirtschaftliche Bedeutung. Umsatzsteigerungen, Verluste oder zu erwartende Schäden hängen vom Wetter ab. Ein Großteil der Wirtschaft ist derzeit wetterabhängig, dennoch nutzen immer noch verhältnismäßig wenig Unternehmen die Möglichkeiten von Wetterinformationen strategisch zu ihrem Vorteil.

Fast alle Bereiche der Wertschöpfungskette werden vom Wetter beeinflusst, so beispielsweise beim Transport von Waren oder beim Angebot in Kaufhäusern. Versicherungsunternehmen behalten die Wetterprognosen und -entwicklungen im Blick, um ihre Risikomodelle anzupassen, oder beispielsweise Stromversorger, um durch schlechtes Wetter erhöhte Bedarfe frühzeitig zu prognostizieren oder durch schlechtes Wetter die geringere Ausbeute von Sonnenkollektoren vorherzusagen.

Das Wetter stellt somit für viele Unternehmen einen der größten externen Einflussfaktoren auf das Geschäftsergebnis dar, unabhängig von der jeweiligen Branche (Handel, Landwirtschaft, Transport und Logistik, Tourismus, Versicherungen, Energiewirtschaft usw.).

In der Regel wird jedoch nur reaktiv auf die Wetterentwicklung reagiert. Dabei ergeben sich heute neue Möglichkeiten, proaktiv mit Wetterinformationen umzugehen. Der Einsatz moderner Sensorik, genauer Wetterprognosen und Analyseverfahren sowie die Nutzung mobiler Geräte eröffnen neue Optionen für die Unternehmensplanung. Die Verknüpfung mit Unternehmensdaten, externen Datenquellen und kognitiven Werkzeugen schafft zusätzliche Potenziale. Die Zusammenführung aller Informationen im Rahmen von BIA-Plattformen, die die Verknüpfung ermöglichen, führt zu zielsicheren Vorhersagen für die jeweiligen Geschäftsprozesse und deren Steuerung.

Genau an diesem Punkt setzen Dienstleister wie »The Weather Company« an. Das Unternehmen ist einer der weltweit größten Wetterdienstleister und stellt seinen Kunden präzise Wetterdaten zur Verfügung, um damit besser auf Wetterphänomene reagieren zu können [Weather].

### **Social-Media-Daten**

Social-Media-Daten sind Informationen, die von individuellen Benutzern erstellt und veröffentlicht und von allgemein zugänglichen Seiten eingesammelt werden. Dazu gehören z. B. soziale Netzwerke wie Twitter, Facebook, Instagram oder Meinungs- und Bewertungsseiten sowie Multimedia-Plattformen wie z. B. YouTube.

Social Media ist ein Medium, das Internetnutzern aus aller Welt dient, Informationen in Echtzeit erstellen und teilen zu können. Dies ist sowohl der größte Vorteil als auch die größte Hürde für das Potenzial von Social-Media-Daten. Zwar stehen uns sofort zugängliche, reale Informationen, die in Echtzeit von Menschen produziert werden, zur Verfügung, um aber sinnvolle Schlussfolgerungen aus diesen Daten zu ziehen, müssen wir die richtigen Parameter für die Nutzung bestimmen.

Grundsätzlich stellen sich im Zusammenhang mit Social Media die Fragen, welchen Nutzen die Daten geschäftlich haben und wie der Zugriff erfolgen kann. Dabei ist zwischen Social-Media-Daten und Social Media Intelligence zu unterscheiden. Mit Social-Media-Daten bezeichnet man das Rohmaterial. Dieses wird mit bestimmten Zielen und unter Anwendung von Kriterien gesammelt und ausgewertet. Dies wird dann ganzheitlich als Social Media Intelligence verstanden.

Die Zielsetzung von Social Media Intelligence kann dabei sehr unterschiedlich sein, beispielsweise die Aufnahme des Stimmungsbilds am Markt zu den angebotenen Produkten, die Verbesserung des Kundenservice oder die Sammlung von Feedback zu Marketingkampagnen.

Um ein Gefühl zu erhalten, wie ein interessierendes Produkt am Markt wahrgenommen wird, können Wortmeldungen aus sozialen Netzwerken wie z. B. Twitter gesammelt werden. Werkzeuge erlauben eine Suche nach relevanten Hashtags unter Verwendung der Produktnamen. Geliefert werden alle Tweets, die wiederum mithilfe einer Sentimentanalyse bewertet werden können. Dabei



wird ein Wert ermittelt, inwieweit die Aussage positiv oder eher negativ ist. Dies dient als Grundlage für eine Stimmungsanalyse hinsichtlich der Marke und auch um zu sehen, wie hoch der Prozentsatz an Beschwerden, Anfragen oder positivem Feedback ist. Die Tweets können für tiefergehende Analysen gespeichert werden.

Jedoch ist auch bei der Nutzung von Social-Media-Daten Vorsicht geboten: So kann nicht immer davon ausgegangen werden, dass die in sozialen Netzwerken geäußerte Meinung die am weitesten verbreitete ist. Es existiert auch eine große Nutzergruppe von »Schweigern« und Nicht-Nutzern von Social Media. Des Weiteren werden die jeweiligen sozialen Netzwerke von unterschiedlichen Nutzergruppen bevorzugt (alt, jung, Frau, Mann usw.). Darüber hinaus existiert, auch wenn von den Anbietern nicht gewünscht, eine Vielzahl von Fake-Accounts. All dies kann das Bild stark verzerren.

Der wohl größte Vorteil von Social Media ist das enorme Volumen an Informationen über Demografien und menschliche Verhaltensmuster weltweit. Diese Informationsquelle lässt sich sofort durch die Werbepattformen der Social-Media-Anbieter nutzen, auch ohne fragwürdige Techniken einzusetzen, die die Daten von Einzelpersonen laden. Letztere müssen stets unter Gesichtspunkten des Datenschutzes bewertet und im praktischen Einsatz auf öffentliche Daten begrenzt bleiben.

## **Open Data**

Open Data bezeichnet öffentlich publizierte Daten, die oft mit weniger beschränkten Lizenzfreigaben und in einfach lesbaren Formaten angeboten werden. Die Bereitstellung erfolgt in der Regel online, wodurch die Wiederverwendbarkeit und der freie Zugriff leicht ermöglicht und gefördert wird [Lucke & Geiger 2010].

Zu Open Data gehören Daten öffentlicher oder privater Herkunft, beispielsweise Lehrmaterial, Geodaten, statistische Wetterdaten, ausgewählte Finanzinformationen (Kurse, Risiken, Prognosen), Statistiken, Verkehrsinformationen, Forschungsergebnisse sowie Hörfunk- und Fernsehsendungen. Dabei kann es sich um Datenbestände staatlicher Stellen oder privatwirtschaftlich agierender Unternehmen, Hochschulen sowie Non-Profit-Einrichtungen handeln. Zu beachten ist jedoch stets die Lizenzbedingung der im Einzelnen betrachteten Quelle.

Es existieren mittlerweile eine Vielzahl von Open-Data-Angeboten am Markt, die in sehr unterschiedlicher Form auch für Unternehmen nutzbar sind. Im Folgenden werden einige Beispiele vorgestellt.

### ■ **International Monetary Fund**

Der IMF wurde 1945 gegründet und besteht derzeit aus 189 Ländern, die sich für die Förderung der weltweiten währungspolitischen Zusammenarbeit, die Sicherung der Finanzstabilität, die Erleichterung des internationalen Handels, die Förderung von hoher Beschäftigung und nachhaltigem Wirtschaftswachstum sowie die Verringerung der Armut in der Welt einsetzt.

Der IMF dient als globaler Knotenpunkt für Wissen in wirtschaftlichen und finanziellen Fragen. Auf der Grundlage seiner Erfahrungen aus verschiedenen Ländern in unterschiedlichen Entwicklungsstadien teilt der IMF dieses Wissen mit den Mitgliedstaaten durch praktische Beratung, Schulungen und Peer-to-Peer-Lernen.

Die Kapazitätsentwicklung des IMF ist Teil seines Kernauftrags – neben der Verfolgung der globalen wirtschaftlichen Entwicklungen und der Kreditvergabe an Länder mit Zahlungsbilanzkrisen. Diese Wissensaustauscharbeit konzentriert sich auf die Bereiche öffentliche Finanzen, Geld- und Finanzsysteme, Rechtsrahmen, Statistiken und makroökonomische Rahmenbedingungen.

Der IMF stellt diese Informationen im Rahmen einer Data-Plattform bereit, die es Nutzern ermöglicht, Finanzinformationen weltweit abzufragen [IMF].

#### ■ **Weltbank-Datenbank**

Die Weltbankgruppe besteht aus fünf Organisationen, der International Bank for Reconstruction and Development, der International Development Association, der International Finance Corporation, der Multilateral Investment Guarantee Agency und dem International Centre for Settlement of Investment Disputes.

Sie unterstützen Entwicklungsländer durch politische Beratung, Forschung und Analyse sowie technische Hilfe. Die analytische Arbeit bildet die Grundlage für die Finanzierung durch die Weltbank und trägt dazu bei, die Entwicklungsländer über Investitionen zu informieren. Darüber hinaus unterstützen sie den Kapazitätsaufbau in den betreuten Ländern.

Die Weltbank stellt eine Open-Data-Plattform bereit, die einen freien Zugang zu globalen Entwicklungsdaten ermöglicht. Die Dienste bestehen u. a. aus dem Open-Data-Katalog, weltweiten Finanzdaten, weltweite Entwicklungsindikatoren, erweiterte Tools zum Exportieren, Herunterladen und Analysieren der Daten sowie zum Erstellen benutzerdefinierter Visualisierungen und gefilterter Ansichten [WB].

Die Daten können abgerufen werden und als Grundlage für weitere Anwendungen dienen, mit denen Entwickler und Analysten spezielle Fragestellungen zum Beispiel im strategischen Marketing und in der Vertriebsplanung beantworten können. So sind etwa die Absatzerwartungen im Investitionsgüterumfeld stark von der globalen wirtschaftlichen Entwicklung abhängig.

#### ■ **Open-Government-Plattformen**

Open-Government-Plattformen ermöglichen u. a. Bürgern die Handlungen von Abgeordneten nachzuvollziehen. Neben Informationen, wie jeder Abgeordnete abgestimmt hat, kann auch nachvollzogen werden, welcher Abgeordnete wie viel Spenden erhält. Hierdurch kann z. B. Korruption und Stimmenkauf durch Lobbyisten vorgebeugt werden.

Die Daten können dazu genutzt werden, im Rahmen von Netzwerkanalysen hinsichtlich Korruptionsbekämpfung zu unterstützen.

### ■ Statistisches Bundesamt

Das Statistische Bundesamt liefert qualitativ hochwertige statistische Informationen über Deutschland. Dazu gehören Indikatoren zu Konjunktur, Nachhaltigkeit, Globalisierung und Bürokratiekosten, Informationen zur Preis und Gehaltsentwicklung u. v. m.

Mit GENESIS-Online stellt das Statistische Bundesamt eine Datenbank zur Verfügung, die ein breites Themenspektrum fachlich gegliederter Ergebnisse der amtlichen Statistiken enthält.

Die Daten können durch zeitliche, sachliche und regionale Auswahlmöglichkeiten gefiltert und abgerufen werden. Die Ergebnisse können in verschiedenen Formaten (Excel, CSV, HTML) abgespeichert werden.

Darüber hinaus besteht die Möglichkeit, die Daten über einen Webservice abzurufen und über eine automatisierte Verarbeitung in bestehende Datenmanagementprozesse eines Unternehmens zu überführen [Destatis 2018].

Die Daten können somit den BIA-Prozessen in den Unternehmen zugänglich gemacht und in Warehouses integriert werden.

### ■ Paradise Papers

Die Paradise Papers stellen in Tausenden Fällen dar, wie Einzelpersonen und Unternehmen auf dem Wege von Offshore-Gesellschaften Steuervermeidung betreiben.

Die geleakten Unterlagen enthalten Informationen zu mehr als 120 Staats- und Regierungschefs sowie Politikern aus 47 Ländern. Diese Daten können in die internen Analysen von Unternehmen und Sicherheitsbehörden einfließen, um diese Netzwerke zu erkennen, zu bekämpfen oder Geschäfte gezielt mit diesen Unternehmen zu vermeiden [ICIJ].

Für die Nutzung entstehen derzeit Open-Data-Anwendungen und -Plattformen, die es ermöglichen, Daten zu unterschiedlichen Themen bereitzustellen. Diese können dann in eigene Datenhaushalte gespeichert, historisiert und für Auswertungen aufbereitet werden. Da sich diese Daten außerhalb des Unternehmens befinden und sie keine sensiblen Unternehmensdaten enthalten, können diese Datenhaltungen auch problemlos in der Cloud verwaltet werden. Über die Möglichkeiten von hybriden Cloud-Szenarien können die Daten dann für Analysezwecke mit internen Daten verknüpft werden. Eine Integration eines Open Data Hub in die Unternehmensarchitektur ist somit nicht zwingend erforderlich.

## 1.3 Organisationsformen der Cloud

Neben den technischen Servicemodellen werden die Organisationsformen der Cloud-Bereitstellung unterschieden. Die Definitionen für Begriffe wie z.B. »private« und »public« wurden dabei vom NIST (National Institute of Standards and Technology) geprägt und sind seit 2014 auch in ISO/IEC 17788 verfügbar

[ISO/IEC 2014]: NIST 800-145 listet fünf Charakteristika für Cloud Computing [Mell & Grance 2011]:

- Selbstzuweisung von Leistungen
- Zugreifbarkeit über Netze für vielfältige Endgeräte
- Ressourcen-Pooling und Multimandanten-Fähigkeit
- Flexible Bereitstellung und Freigabe von Ressourcen
- Messbarkeit und automatische Kontrolle und Optimierung

Diese Leistungsmerkmale können über unterschiedliche Liefermodelle bereitgestellt werden. NIST definiert vier Liefermodelle.

### 1.3.1 Liefermodelle der Cloud

#### Public Cloud

Die Public Cloud bietet den Zugang zu einer multimandantenfähigen IT-Infrastruktur für die öffentliche Nutzung über das Internet. In der Public Cloud können Kunden IT-Infrastruktur auf einer flexiblen Basis des Bezahlens (monatliche Abonnements, leistungsabhängige Abrechnung) mieten, ohne in Rechner- und IT-Infrastruktur investieren zu müssen.

BI & Analytics profitiert von Public Clouds durch die Bereitstellung von temporären Infrastrukturen oder Plattformen. Dies ist stets hilfreich, wenn Proof of Concepts durchgeführt werden müssen, um innovative Konzepte zu testen. Darüber hinaus ist stets zu erwägen, ob Teile der BI & Analytics dauerhaft von Leistungsmerkmalen der Cloud profitieren können (z.B. Skalierbarkeit, Flexibilität).

#### Private Cloud

Die Private Cloud bietet eine virtualisierte IT-Infrastruktur, bei der sich die Rechner innerhalb der eigenen Organisation (Behörde, Firma, Verein) befinden. Im Rahmen der Private Cloud kann zwischen zwei weiteren Betriebskonzepten unterschieden werden: Bei der Managed Private Cloud erfolgt der Betrieb durch den externen IT-Dienstleister. Die Cloud-Infrastruktur verbleibt in der Regel im Hause des Kunden und in seinem Eigentum. Bei der Outsourced Private Cloud übernimmt ein externer IT-Dienstleister von einem Kunden eine dedizierte Cloud-Infrastruktur (oder baut sie auf) und betreibt diese vollverantwortlich. Die Infrastruktur steht physisch beim externen Partner, der auch Eigentümer der Assets ist.

Private Clouds sind ein Mittel zur Flexibilitätssteigerung der internen IT und heute weitgehender Standard. Hier ergeben sich für BI & Analytics keine besonderen Schlussfolgerungen.

## Hybrid Cloud

Die Hybrid Cloud kombiniert Public und Private Clouds nach den Bedürfnissen ihrer Nutzer.

Der typische Ansatzpunkt für Hybrid Clouds im BIA-Umfeld ist der Aufbau eines Data Warehouse, in dessen Architektur Cloud-Komponenten integriert werden. Beispiel hierfür ist die Integration von Datenquellen, die über Cloud-Services günstig angebunden werden können wie Social-Media- oder Wetterdaten. Diese lassen sich auch durch Microservices sehr gezielt aus der Cloud abrufen.

## Community Cloud

Die Community Cloud entspricht im Wesentlichen der Public Cloud jedoch für einen ausgewählten Nutzerkreis, der räumlich verteilt ist und sich die Kosten teilt. Hierunter fallen oft städtische Behörden, Universitäten, Forschungsgemeinschaften, Genossenschaften, Kooperationsnetze von Unternehmen usw.

Der naheliegende Anwendungsfall für Community Clouds im Bereich BI & Analytics findet sich immer dann, wenn vertrauliche Daten in ähnlicher Form bei unterschiedlichen Anwenderunternehmen vorliegen und leicht über die Cloud gesammelt werden können. Um diese übergreifend auszuwerten, wird ein vertrauenswürdiger Partner etabliert, der die Daten zusammenführt und Erkenntnisse an die Mitglieder der Community verteilt. Praktisch ist dies etwa im Bereich Industrie 4.0 anzutreffen: Ein Maschinenbauunternehmen tritt hier nicht selten als vertrauenswürdiger Partner auf und sammelt Prozessdaten aus der Fertigung der angeschlossenen Produktionsunternehmen für Zwecke der prädiktiven Wartung.

### 1.3.2 BIA-Cloud-Strategie

Cloud-basierte BIA-Tools und -Services sind in den letzten Jahren immer reifer und beliebter geworden. Vor allem kleine und mittlere Unternehmen haben erkannt, dass das Outsourcing von BIA-Infrastrukturen Ressourcen bereitstellen kann, die bisher nur großen Unternehmen zur Verfügung standen. Nun sind die BIA-Anwendungen in kleinen und mittleren Unternehmen oft überschaubar hinsichtlich der genutzten Komponenten und der benötigten BIA-Szenarien. Operative Systeme sind oft schon Cloud-basiert und im BIA-Umfeld werden nur die Daten genutzt, die ohnehin schon in der Cloud sind. In diesem Fall liegen die Vorteile einer Cloud-basierten BIA-Lösung nahe.

In Unternehmen, in denen die BIA-Landschaften komplexer und mitunter historisch gewachsen sind und operative Systeme ausschließlich oder größtenteils als On-Premises-Systeme realisiert sind, ist die Entscheidung nicht mehr ganz so einfach. Die Komplexität der Architektur, im Besonderen hinsichtlich der Abhängigkeiten, sowie eventuell gesetzliche Regelungen und Anforderungen von Aufsichtsbehörden erschweren eine Migration in die Cloud. Kriterien sind festzulegen, die regeln, unter welchen Bedingungen ein Einsatz in der Cloud machbar ist.

Der Ausbau der BIA-Landschaft um Anwendungen für die Verwertung von Open Data ist sicher ein Szenario, das für den Aufbau in der Cloud geeignet ist. Aber auch die Nutzung von Analytic Services spricht für die Cloud. So können z.B. Analytic Labs in der Cloud schnell bereitgestellt, genutzt und nach Gebrauch wieder abgebaut werden.

Wie bereits erwähnt bieten Softwarehersteller ihre Reporting & Dashboarding-Lösung mehr und mehr in der Cloud an. So kann man die Möglichkeit in Betracht ziehen, auch das Reporting in die Cloud auszulagern. Wesentlich dabei ist jedoch, welche Daten hier mit übertragen werden. Dies sollte bei der Werkzeugauswahl geprüft werden.

Ein typischer Ansatzpunkt für den Einsatz von Cloud-BIA ist natürlich die Durchführung von PoCs und prototypischen Entwicklungen auf Basis von Sandboxen. Dies stellt immer wieder Anforderungen an die Basis-IT-Einheiten, hierfür Server bereitzustellen und diese für den Zeitraum zu betreiben. Die Nutzung von Cloud-Infrastrukturen kann hierbei helfen, die Aufwände und Kosten zu reduzieren. Die Nutzung von vorkonfigurierten Services ermöglicht auch die Erarbeitung von prototypischen Lösungen, um Sachverhalte und Abhängigkeiten besser zu verstehen.

Es gibt somit keine für alle Unternehmen einheitlich gültige Cloud-Strategie. Es zeichnet sich aber ab, dass die entstehenden neuen BIA-Landschaften eine Kombination aus On-Premises-Anwendungen und IaaS, PaaS oder SaaS sein werden. In manchen Fällen werden sicher auch DWHaaS-Angebote in Betracht gezogen werden, um z.B. Data Marts in der Cloud zu publizieren und diese mit weiteren Services zu verknüpfen.

Wesentlich für die Entscheidung wird die Sensibilität der Daten, die Anforderungen an die Flexibilität und die Möglichkeiten der Kostenreduktion sein. Der Schritt für BIA in die Cloud wird in den Unternehmen daher schrittweise zu gestalten sein, basierend auf hybriden Ansätzen.

## **1.4 Nutzen und Risiken**

Cloud Computing bietet heute für BI & Analytics unübersehbare Vorteile. Die Risiken sind beherrschbar und sollten nicht zur Begründung einer pauschalen Vermeidungshaltung dienen.

### **1.4.1 Vorteile der Cloud**

Wesentliche Argumente für die Nutzung von Cloud Computing sind die Verlagerung der Investitionen für Hard- und Software sowie die Variabilisierung der Betriebskosten, abhängig von der tatsächlichen Nutzung der Systeme. Sprungfixe Kosten für Kauf und Einrichtung der Infrastrukturen entfallen. Auch kann eine Cloud-basierte Infrastruktur schnell wieder außer Dienst gestellt werden. Dies

vereinfacht die Entscheidungsprozesse hinsichtlich der Anschaffung und Bereitstellung von Diensten aufseiten des Anwenders.

Weitere Vorteile ergeben sich bei stark schwankender Nachfrage. Cloud Computing ermöglicht, die verfügbare Kapazität kurzfristig an den tatsächlichen Bedarf anzupassen.

Vor eine pauschalen Erwartung, Cloud-Services seien unschlagbar günstig, ist jedoch zu warnen. Durch die Buchung eines Cloud-Service schließt der Anwender einen komplexen Servicevertrag ab, der umfangreiche IT-Leistungen auf die Seite des Cloud-Service-Providers verlagert. Diese Leistungen werden auch bei Berücksichtigung einer Kostendegression durch Skaleneffekte stets ihren Preis haben. Positiv ist anzumerken, dass führende Cloud-Provider heute gut kalkulierbare Preismodelle und transparente, servicebasierte Abrechnungen bieten.

#### 1.4.2 Skepsis gegenüber der Cloud-Sicherheit

Mit dem Einsatz von Cloud-Technologien geben Unternehmen die Wartung und teilweise auch den Betrieb in die Hände der Cloud-Service-Anbieter, und damit auch die Aufgaben hinsichtlich IT- und Datensicherheit.

Derzeit sind Sicherheitsbedenken wohl die häufigsten Argumente der Cloud-Verweigerer. Die neuen gesetzlichen Regelungen der Datenschutz-Grundverordnung und die Anpassungen des Bundesdatenschutzgesetzes führen dazu, dass die Unternehmen eher zögerlich hinsichtlich Cloud-basierter Lösungen sind. Davon betroffen ist dann, aus verständlichen Gründen, die Auslagerung des Data Warehouse in die Cloud.

Zumal es Rechtsstreitigkeiten großer US-Anbieter mit der US-Regierung gibt, dass Staaten wie die USA weiterhin unter bestimmten sicherheitsrelevanten Bedingungen auf Daten von US-Firmen zugreifen wollen, unabhängig in welchem Land die Daten gespeichert sind. Das Thema Datenschutz und Datensicherheit bleibt damit ein wesentlicher Faktor hinsichtlich der Cloud-Nutzung für Data Warehouses in den Unternehmen.

Darüber hinaus steht die klassische IT-Governance, die sich über die IT-Betriebshoheit (Installationsrechte, Netzzugang etc.) definiert, vor einer völlig neuen Herausforderung: Cloud-BIA-Services werden oft vorbei an allen üblichen Freigabeprozessen direkt über den Webbrowser implementiert – vom Datenbank-Setup über die Datenbeladung bis hin zu Frontend-Entwicklung und Rollout. Hierfür müssen explizite Richtlinien erarbeitet werden. Unterbleibt dies, ist es nicht verwunderlich, wenn Fachbereiche mit fehlendem Problembewusstsein die verführerische Einfachheit von Cloud-BIA nutzen und so vertrauliche Daten das Unternehmen verlassen.

### 1.4.3 Performance-Zusagen in der Cloud

Oft wird die Performance der Cloud in Verbindung mit dem Transport von Massendaten betrachtet. Im Netzausbau mag Deutschland noch Potenzial haben, was sich auch auf die Übertragungsgeschwindigkeiten von Daten in die Cloud auswirkt. Jedoch werden Provider auch in Zukunft hier eher keine Übertragungsleistung garantieren, da nicht ermittelt werden kann, wo der Performance-Verlust entsteht (Server, Leitung, die eigene Rechnerleistung usw.). Dem Kunden wird in der Regel eine virtuelle Maschine bereitgestellt, der Prozessoren, RAM und Storage zugewiesen werden. Diese können aber von mehreren Rechnern genutzt werden. Lediglich die Nutzung von Bare-Metal-Machines ermöglicht die Einhaltung von Leistungsparametern. Diese Angebote sind zuzeit jedoch seltener und entsprechend kostspielig.

## 1.5 Fazit

Für flexible und schnelle Hilfe bei Analysen (Sandboxes, Analytic Labs) oder für Testzwecke (z.B PoCs) sind Cloud- und DWHaaS-Angebote schon heute sehr gut geeignet, weil keine eigene kostenintensive Infrastruktur aufgebaut werden muss. Auch bei weniger kritischen Daten, die u.U. ohnehin in der Cloud liegen, ist Cloud-BIA eine natürliche Wahl.

Die zu erwartende Änderung im Umgang mit Daten sowie die abzuwartende Klärung hinsichtlich der Belastbarkeit der Rechtslage rund um die Cloud werden Zeit benötigen. Diese ist notwendig, um das Vertrauen in Cloud-basierte BIA-Plattformen zu schaffen und Unternehmen zu bewegen, ihre sensiblen, erfolgsrelevanten Daten in die Cloud zu verlagern.

Bis dahin ist Anwenderunternehmen zu empfehlen, mit ausgewählten Cloud-Anbietern eine strategische Zusammenarbeit zu etablieren. Diese erlaubt eine explizite Ausrichtung aller sicherheitsrelevanter Aspekte in der IT-Infrastruktur in enger Abstimmung mit den strategischen Cloud-Anbietern.