

1 Einführung in die BIA-Architekturen


Peter Gluchowski • Frank Leisten • Gero Presser

Der vorliegende Beitrag setzt sich das Ziel, die Rahmenbedingungen für komplexe Business Intelligence & Analytics-(BIA-)Landschaften zu beleuchten. Den Ausgangspunkt für die Betrachtungen bildet der folgende Abschnitt, der BIA-Trends und -Entwicklungen in der letzten Dekade punktuell aufgreift und die Bedeutung für die zugehörigen dis-positiven Ökosysteme herausarbeitet. Danach erfolgen eine Abgrenzung und Einordnung der BIA-Architektur zu verwandten Themen wie Unternehmensarchitektur, IT-Architektur, Anwendungsarchitektur und Infrastruktur (Abschnitt 1.2). Anschließend nähert sich Abschnitt 1.3 dem Architekturthema aus einer Datenperspektive, indem die Datenstrategie, die Wertermittlung von Daten und das Datenmanagement im Vordergrund der Betrachtung stehen. Der anschließende Abschnitt 1.4 beleuchtet die Anforderungen an eine ganzheitliche BIA-Architektur aus der Perspektive unterschiedlicher Anspruchsgruppen und macht deutlich, dass sich die Vorstellungen und Ziele erheblich voneinander unterscheiden können. Schließlich greift Abschnitt 1.5 die klassische Hub-and-Spoke-Architektur und die Schichtenarchitektur für BIA-Ökosysteme auf und verweist auf die zugehörigen Defizite.

1.1 BIA-Trends und -Entwicklungen

In der letzten Dekade lässt sich eine zunehmende Komplexität analytischer Architekturen feststellen. Waren es noch vor zehn Jahren die klassischen Data-Warehouse-zentrierten Architekturkonzepte, die fast flächendeckend und ausschließlich Verwendung fanden, haben in der Zwischenzeit vielfältige zusätzliche Komponenten und Technologien Einzug in die BIA-Landschaften der Unternehmen gehalten.

Unterstützt wurde diese Entwicklung nicht zuletzt durch die intensive Diskussion um Big Data, die durch die Hypothese geleitet ist, dass die herkömmlichen Konzepte und Technologien nicht dazu in der Lage sind, alle aktuellen Anforderungen in geeigneter Form zu erfüllen. So greifen einige Veröffentlichungen zu dem Thema auf eine Negativabgrenzung zurück und stellen heraus, dass Big Data

Diese Leseprobe haben Sie beim
 edv buchversand.de heruntergeladen.
Das Buch können Sie online in unserem
Shop bestellen.
[Hier zum Shop](#)

dann gegeben ist, wenn die Kapazitäten und Funktionalitäten der klassischen Datenhaltung, -aufbereitung und -auswertung sich als nicht ausreichend erweisen [Dittmar et al. 2016, S. 3]. Zumeist wird Big Data heute durch die charakteristischen Eigenschaften beschrieben. Dann zeichnet sich Big Data nicht allein durch das immense Datenvolumen (Volume) aus, sondern ebenso durch die erhebliche Vielfalt an Datenformaten (Variety) sowie durch die Geschwindigkeit (Velocity), mit der neue Daten entstehen sowie verfügbar und damit analysierbar sind [Eaton et al. 2012, S. 5].

Allerdings lassen sich zahlreiche weitere Begrifflichkeiten mit dem Anfangsbuchstaben »V« und somit weitere Dimensionen identifizieren, mit denen Big Data umschrieben wird. Beispielsweise adressiert Veracity als Wahrhaftigkeit oder Richtigkeit der Daten eine weitere Eigenschaft von Big Data, zumal Auswertungen und die damit verbundenen Entscheidungen hierauf beruhen und falsche Daten zu fehlerhaften Analyseergebnissen führen können. Aufgrund der Datenvielfalt und des Datenvolumens erweist sich eine Überprüfung der Daten jedoch häufig als schwierig [Klein et al. 2013, S. 321]. Als weitere Begrifflichkeiten mit »V« finden sich Validity, Volatility, Variability und vor allem Value, auf die hier allerdings nicht weiter eingegangen wird [Gandomi & Haider 2015, S. 139; Khan et al. 2014, S. 3]. Es liegt auf der Hand, dass hieraus gänzlich neue Bedarfe resultieren, die es zu erfüllen gilt.

Auch seitens der Datenanalyse haben sich in den letzten Jahren bemerkenswerte Veränderungen eingestellt, die sich in einer verstärkten Hinwendung zu anspruchsvollen statistisch-mathematischen Verfahren unter Oberbegriffen wie künstliche Intelligenz, Machine Learning oder Data Science zeigen. Derzeit erweisen sich vor allem komplexe künstliche neuronale Netze (Deep Learning) als leistungsfähig, mit denen die Erforschung von Strukturzusammenhängen (Datenmustern) in Datenbeständen eine neue Qualität erreicht [Dorer 2019, S. 119 ff.].

Als Konsequenz aus diesen Entwicklungen erfolgte in zahlreichen Unternehmen eine zumindest teilweise Abkehr beispielsweise von den klassischen, festplattenorientierten relationalen Datenbanksystemen hin zur schemalosen und verteilten Ablage des Datenmaterials, mit der sich auch große und polystrukturierte Inhalte organisieren lassen. Daneben mündet die Forderung nach hoher Verarbeitungsgeschwindigkeit in neuen Herausforderungen, die eine Erweiterung oder Ergänzung der bislang üblichen Batch-orientierten Aufbereitung des Datenmaterials für analytische Zwecke zur Folge hat – spätestens dann, wenn Datenströme zu verarbeiten sind.

Begünstigt wird die Veränderung durch zahlreiche neue Technologien. Bezogen auf die Speicherung von Daten sei hier etwa auf In-Memory-Konzepte, NoSQL-Datenbanksysteme (z. B. als Key-Value Store) sowie auf Cloud-Technologien verwiesen. Im Frontend-Sektor dagegen haben Self-Service-Werkzeuge breiten Raum eingenommen.

Auch aus organisatorischen Gründen haben sich im letzten Jahrzehnt die Voraussetzungen für die Gestaltung von BIA-Architekturen geändert. So erfor-

dert die zunehmende Hinwendung zu agilen Gestaltungsmethodiken mit kurzen Entwicklungszyklen, dass sich inkrementelle und iterative Veränderungen im Systemaufbau auch mit den vorhandenen Landschaften realisieren lassen. Aufgrund des engen zeitlichen Rahmens erweist es sich dabei teilweise als unumgänglich, dass einzelne Entwicklungsschritte durch Automatisierungsverfahren und -komponenten beschleunigt werden. Aber auch aus dem Betrieb von BIA-Lösungen ergeben sich Beschleunigungsbedarfe, die oftmals unter dem Begriffsgebilde DataOps diskutiert werden [Detemple 2020]. Gefordert wird hier sowohl eine Datenpipeline als auch eine Analytics-Pipeline zur möglichst zeitnahen Zurverfügungstellung von Berichten, Dashboards und Analytics-Modellen für den Endanwender.

Weitere Rahmenbedingungen für die BIA-Landschaft ergeben sich aus externen, regulatorischen, aber auch internen Vorgaben, die es zu erfüllen gilt. Als wichtige regulatorische Vorgabe lässt sich die Datenschutz-Grundverordnung (DSGVO) anführen, aus der sich die Notwendigkeit eines besonderen Umgangs mit personenbezogenen Daten und der architektonischen Umsetzung ableiten lässt. In einzelnen Branchen existieren darüber hinaus spezielle Regularien, die weit über den einfachen gesetzlichen Standard hinausreichen. So kann für den Finanzdienstleistungssektor das Regelwerk der BCBS 239 angeführt werden, aus dem sich weitreichende Anforderungen an die Transparenz und Nachverfolgbarkeit der Verarbeitung von Daten ergeben.

Vor diesem Hintergrund wird deutlich, dass eine einfache Architektur mit wenigen Komponenten heute kaum ausreichen kann, um allen Anforderungen gerecht zu werden. Vielmehr stellt sich die Aufgabe, ein analytisches Ökosystem zu gestalten, in dem jeder Baustein definierte Funktionen übernimmt und dabei seine spezifischen Stärken einbringt. Naturgemäß ergibt sich hieraus die steigende Komplexität der Gesamtlandschaft, zumal das reibungslose Zusammenspiel der einzelnen Komponenten eine Herausforderung darstellt.

1.2 Architekturkonzepte und -facetten

Der Begriff Architektur findet in zahlreichen Wissensdisziplinen und thematischen Bereichen Verwendung. Allgemein repräsentiert eine Architektur die Gesamtheit aller beschreibenden Darstellungen (Entwurfsartefakte) der erkenntnisrelevanten Objekte derart, dass diese den Anforderungen entsprechend produziert und betrieben werden können (Qualität). Idealerweise bleiben die grundlegenden Teile der Beschreibung möglichst unverändert über die Nutzungsdauer erhalten [Zachman 1997], können aber an geänderte Bedingungen angepasst werden. Die Artefakte bilden neben der Repräsentation von Objekten auch deren Funktionen, Schnittstellen und Beziehungen sowie dynamische Aspekte ab, wie den zeitlichen Ablauf von Austauschbeziehungen [Krcmar 2015, S. 280 f.].

Im Kontext von Informationssystemen umfasst dies die modellhafte Beschreibung der grundsätzlichen Struktur eines Systems mit seinen Elementen, der Beziehungen zwischen den Elementen sowie den Beziehungen des Systems zur Umwelt

[ISO 2000; Knoll 2018, S. 889]. Neben der Spezifikation seiner Komponenten und ihrer Beziehungen unter allen relevanten Blickwinkeln lassen sich auch die Konstruktionsregeln zur Erstellung des Bauplans [Sinz 2019] sowie die Prinzipien zur Konstruktion, Weiterentwicklung und Nutzung des Systems zu einer Informationssystem-Architektur zählen [IEEE 2000].

Durch die umfassende, globale Sicht auf ein Informationssystem, die alle relevanten Komponenten beinhaltet, unterscheidet sich die Architektur von eingeschränkteren Ansätzen (z.B. der unternehmensweiten Datenmodellierung). Zudem erfolgt die Konzentration auf eher aggregierte Elemente und Beziehungen, um die Ganzheitlichkeit der Betrachtung zu ermöglichen, ohne den Überblick zu verlieren [Winter & Aier 2019].

Als Teil einer Informationssystem-Architektur beschreibt die Datenarchitektur eines Informationssystems auf Fachkonzept- oder Entwurfsebene die grundlegenden Datenstrukturen und bildet dabei die Datenarchitektur eines ganzen Unternehmens ab oder konzentriert sich als Datenarchitektur eines Anwendungssystems auf einen Ausschnitt des Unternehmens [Winter & Aier 2019]. Demgegenüber repräsentiert die IT-Infrastruktur die technischen Komponenten, bestehend aus Hardware, (System-)Software sowie baulichen Einrichtungen für den Betrieb von (Anwendungs-)Software [Patig et al. 2019].

Um den Bezug zu geschäftlichen bzw. fachlichen Sichtweisen auf die Architekturen und damit ein gutes Business-IT-Alignment zu wahren, sind über die technische Perspektive hinaus weitere Aspekte zu berücksichtigen [Knoll 2018]. So lassen sich strategische und organisationale Ebenen abbilden, die auf den technischen Layern aufsetzen und diese ergänzen.

Auf jeder der betrachteten Architekturebenen finden sich unterschiedliche Objekte, deren Ausgestaltung und Zusammenwirken den Aufbau des Gesamtgebildes bestimmen (vgl. Abb. 1–1). Zur Gestaltung sind verschiedene Modelltypen verwendbar, die beim Entwurf der spezifischen Strukturen unterstützen. So finden sich auf der strategischen Ebene beispielsweise Modelle zur Abbildung von Geschäftsbeziehungen zu Kunden und Lieferanten. Bei der Beschreibung der Organisationsebene gelangen neben Prozesslandkarten und -modellen auch Organigramme sowie (fachliche) Informationslandkarten zur Anwendung. Auf der untersten Ebene, der IT-Infrastrukturebene, finden sich Beschreibungen über das Zusammenspiel (hardwarenaher) technischer Komponenten wie Modelle der Netzwerkinfrastruktur. Die Softwareebene darüber bildet neben den relevanten Datenstrukturen auch den Aufbau der Softwarekomponenten ab, beispielsweise auf Basis von Softwaremodulen oder auch -services.

Eine besondere Rolle spielt bei diesem Konzept die Integrationsebene, die sich als Mittler zwischen betriebswirtschaftlich-fachlicher und technischer Perspektive erweist. Hier werden einzelne Softwarebestandteile zu Anwendungen und Datenstrukturen zu Domänen verknüpft, um einzelne fachliche Prozesse

unterstützen zu können. Infolgedessen lassen sich hier Modelle der Applikationslandschaft und Domänenmodelle verwenden. Neben der Verknüpfungsfunktion erwies sich hier in der Vergangenheit die Entkopplung von fachlichen und technischen Komponenten als hilfreich, um die langsam sich ändernden technischen Gegebenheiten (mit Zykluszeiten von 6 bis 10 Jahren) mit den relativ schnell sich ändernden fachlichen Ebenen (von 3 bis 6 Monaten auf der Organisations-ebene bis zu 1–2 Jahren auf der Strategieebene) zu synchronisieren [Winter 2008, S. 24 ff.].

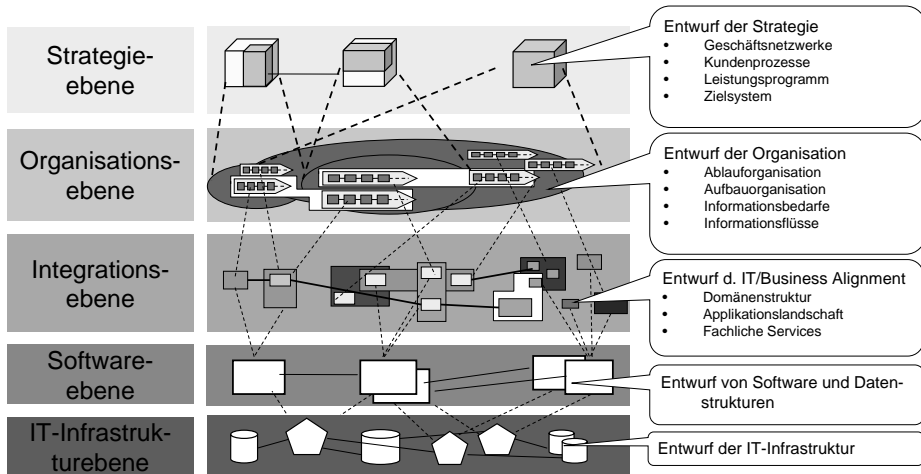


Abb. 1-1 Architekturebenen des Business Engineering [Winter 2010, S. 90]

Vor dem Hintergrund von sich stetig schneller entwickelnden technologischen Innovationen und dem fast flächendeckenden Einzug von agilen Entwicklungsmethoden erweist es sich als fraglich, ob die unterschiedlichen Änderungsgeschwindigkeiten heute noch in dieser Form gegeben sind. Vielmehr scheint es oftmals so, dass die hohe technische Entwicklungsdynamik als Enabler Druck auf die fachlichen Strukturen und Prozesse ausübt. Als Indikator hierfür mag die oft mühselige Suche nach passenden Business Cases gelten, wenn neue, beispielsweise unstrukturierte Datenbestände in den Unternehmen verfügbar sind.

Der vorliegende Sammelband konzentriert sich mit den zugehörigen Beiträgen auf die Integrations- und die Softwareebene. Hierfür sollen unterschiedliche Architekturkonzepte vorgestellt und diskutiert werden, wie sie sich in einzelnen BIA-Ökosystemen präsentieren.

Bevor jedoch auf die konkreten Ausgestaltungen von BIA-Ökosystemen eingegangen wird, sind zunächst die Rahmenbedingungen für eine geeignete Architektur zu beleuchten, die sich sowohl aus strategischen Überlegungen zum Umgang mit Daten als wichtige Ressource als auch aus den konkreten Anforderungen der Stakeholder ergeben.

1.3 Datenbezogene Rahmenbedingungen

Ein Rahmen für die Verarbeitung von Daten in einem Unternehmen besteht im Wesentlichen aus folgenden Teilbereichen: Strategie, Management, Funktion, Prozess und Technologie. Der vorliegende Abschnitt skizziert und positioniert diese Teilbereiche, indem die Handlungsfelder beschrieben und die Wechselwirkungen untereinander aufgezeigt werden. Die folgende Abbildung 1–2 ordnet die Handlungsfelder den entsprechenden Themengebieten zu.

Die strategische Ebene definiert, wie sich Daten im Sinne des Geschäftsmodells nutzen lassen und eine geeignete Datenstrategie (vgl. Abschnitt 1.3.1) abgeleitet werden kann. Die Managementebene schafft ein geeignetes Rahmenwerk zum Umgang mit diesen Daten mittels der Funktionen des Datenmanagements (vgl. Abschnitt 1.3.3). Hierbei werden die Daten zuvor im Rahmen der Data Valuation bewertet und entsprechend ihrer strategischen, wirtschaftlichen und regulatorischen Bedeutung priorisiert und kategorisiert (vgl. Abschnitt 1.3.2). Die Funktionen des Datenmanagements unterstützen oder ermöglichen den Prozess der datenbasierenden Wertschöpfung, die von den Geschäftsfunktionen ausgeführt wird. Datenstrategie und Data Governance regulieren, steuern, verwalten und überwachen die Funktionen des Datenmanagements. Die Management- und Funktionsebenen werden hierbei auf den gesamten Lebenszyklus von Daten angewandt. Die Verwaltung des Lebenszyklus von Daten und die Wertschöpfung auf deren Basis finden im Habitat der Architekturen und ihrer Systemkomponenten statt.

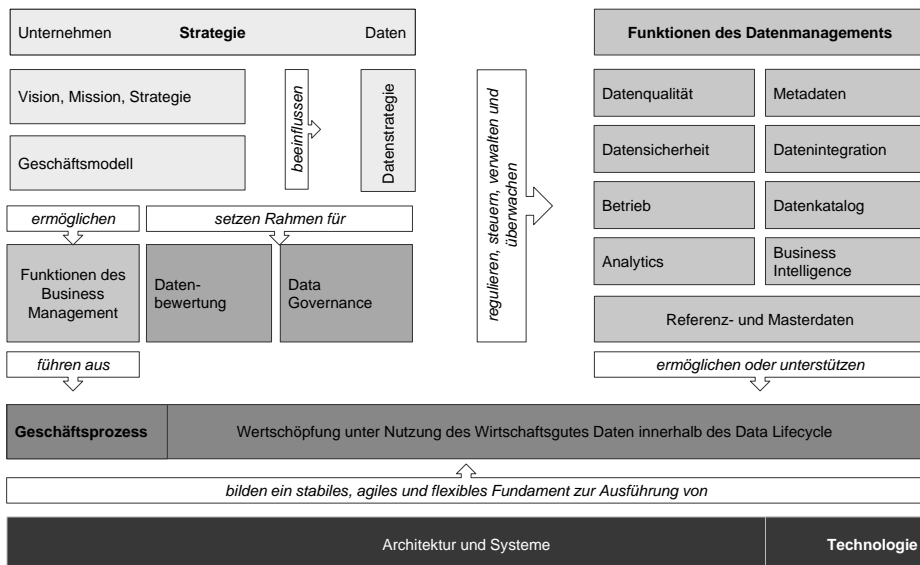


Abb. 1–2 Daten-Ökosystem in einem Unternehmen mit Ebenen und Handlungsfeldern

Dementsprechend fundamental sind die Architekturen und damit die Gesamtheit der Systeme zu gestalten, um eine dauerhafte und nachhaltige Basis zur Wertschöpfung aus Daten als Wirtschaftsgut und deren Verwaltung über den gesamten Lebenszyklus zu gewährleisten. Dabei sind vor allem Stabilität, Agilität und Flexibilität der Architekturen und Systeme sicherzustellen. Nachfolgend wird die Herleitung der Architekturanforderungen, von der Strategieebene ausgehend, beschrieben und in Abschnitt 1.4 als Anforderungen an eine BIA-Architektur zusammengefasst.

1.3.1 Datenstrategie

Allgemein wird auf der strategischen Ebene von der Unternehmensleitung (bzw. von den verantwortlichen Entscheidungsträgern) festgelegt, wie Daten im Sinne der Unternehmung einzusetzen sind. Im Rahmen der Strategiefindung muss die Denkweise über die Bedeutung von Daten an die jeweiligen spezifischen Bedingungen angepasst werden. Zahlreiche Unternehmen setzen Daten nach wie vor ausschließlich zur Unterstützung und Verbesserung bestehender Prozesse im Sinne von Messen und Verwalten ein [Rogers 2017]. Allerdings erfolgt – in immer mehr Organisationen und vor allem in den letzten Jahren – ein Umdenken in Bezug auf die Bedeutung und den Umgang mit Daten, wie in Tabelle 1–1 gegenübergestellt ist.

Früher	Heute
Die Datengenerierung innerhalb eines Unternehmens ist teuer	Daten werden ständig und überall generiert
Das Speichern und Verwalten von Daten stellt eine Herausforderung dar	Die Herausforderung besteht darin, Daten in wertvolle Informationen zu transformieren
Unternehmen nutzen nur strukturierte Daten	Unstrukturierte und semi-strukturierte Daten sind zunehmend nutzbar und stellen einen großen Wert dar
Daten werden in operativen Silos verwaltet	Wert generieren Daten insbesondere durch übergreifende Verbindungen
Daten sind ein Mittel zur Verbesserung von Prozessen	Daten sind ein immaterieller Vermögenswert und dienen damit der Wertschöpfung

Tab. 1–1 Anpassung der strategischen Denkweise [Rogers 2017, S. 139]

Zusammenfassend beinhaltet Tabelle 1–1 folgende Kernaussagen:

- Die Generierung von Daten erfolgt heute ubiquitär von Menschen und Maschinen.
- Die Herausforderung besteht nicht mehr in der Beschaffung von Daten, sondern in der Informationsgewinnung.
- Durch Einsatz fortschrittlicher Technologien und Methoden lassen sich aus Informationen Werte generieren.

Somit stellt sich bei der Strategieentwicklung die Aufgabe, Daten als wesentliche Schlüsselressource und damit wichtiges immaterielles Wirtschaftsgut für die eigene Organisation zu verstehen und zu behandeln, indem ein geeigneter Rahmen zu deren Bewirtschaftung definiert wird (vgl. hier auch die Datenstrategie der UN, abrufbar unter <https://www.un.org/en/content/datastrategy/index.shtml>). Dabei ist nicht zuletzt die zentrale Frage zu beantworten, welche Daten für das jeweilige Geschäftsmodell von Bedeutung sind oder sein könnten. Als exemplarische Einsatzgebiete von Daten, die bei der Definition einer Datenstrategie Bedeutung erlangen können, lassen sich anführen:

- Sammlung heterogener Datenarten für unterschiedlichste Zwecke
- Nutzung von Daten zur Prognose im Rahmen der Entscheidungsfindung
- Nutzung von Daten zur Entwicklung von Produktinnovationen
- Beobachtung des Verhaltens von Kunden
- Kombination von Daten aus diversen Bereichen bzw. Domänen

Aus einer strategischen Perspektive können Daten sowohl eine Supporter-Rolle (Unterstützer) als auch eine Enabler-Rolle (Ermöglicher) einnehmen (vgl. Abb. 1–3).

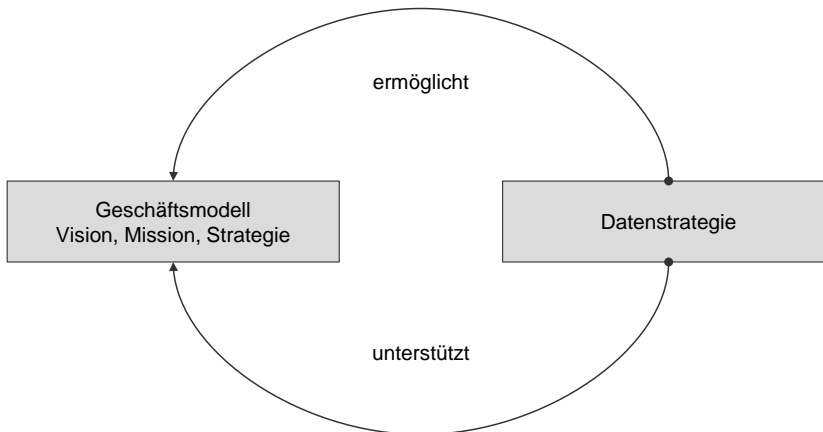


Abb. 1-3 Zusammenhang zwischen Datenstrategie und Geschäftsmodell

Die eingenommene Rolle wird durch den jeweiligen Datenstrategieansatz bestimmt, wobei sich hier defensive von offensiven Ausprägungen abgrenzen lassen. Der defensive Teil verfolgt das Ziel, nachteilige Datenrisiken zu minimieren, und widmet Themen wie Datenschutz, Datenintegrität, Identifizierung, Standardisierung sowie dem operativen Verwalten der Daten besondere Aufmerksamkeit. Im BIA-Kontext wird als Ziel die Bereitstellung einer »Single Source of Truth« bzw. eines »Single Point of Truth« verfolgt. Als Treiber für diese strategische Ausrichtung fungieren u.a. allgemeine Anforderungen an den Betrieb der Lösungen neben regulatorischen Vorgaben, was im Ergebnis zu ausgeprägter Stabilität führt.

Dagegen verfolgt der offensive Ansatz das Ziel, die Wirtschaftlichkeit zu steigern, Performance-Verbesserungen zu erzielen und die Kundenzufriedenheit zu erhöhen. Auf Basis der offensiven Strategie werden »Multiple Versions of Truth« erzeugt. Der offensive Ansatz eröffnet Chancen und erreicht dies durch Anreicherung der Daten und Verwendung analytischer Anwendungen. Insgesamt wird die Agilität und somit auch die Resilienz bzw. Widerstandskraft des Unternehmens verbessert [DalleMule & Davenport 2017].

Bezogen auf die spezifische Situation einer Organisation sind beide Strategien zu beachten und in Betracht zu ziehen, um unter den jeweiligen Rahmenbedingungen (wie Branche und Geschäftsmodell) ein ausgewogenes Verhältnis zwischen Offensive und Defensive im eigenen Haus zu etablieren. Als Basisannahme muss der Datenstrategie die Einsicht zugrunde liegen, dass nur auf Basis einer defensiven Stabilität eine funktionierende Offensive erfolgreich sein kann.

Bei all dem gilt zu beachten, dass Daten als immaterielle Wirtschaftsgüter über andere Eigenschaften als materielle Wirtschaftsgüter verfügen und somit ein individuelles Umfeld zur Bewirtschaftung geschaffen werden muss. Tabelle 1–2 zeigt einige zentrale Unterscheidungskriterien auf.

Materielle Wirtschaftsgüter	Daten als Wirtschaftsgüter
Spezifische Distribution	Einfache Distribution (Internet etc.)
Einfache Ermittlung des Wertes – Bewertung anhand von Marktpreisen möglich	Komplexe und problematische Wertermittlung
Kosten einfach zu ermitteln	Kosten schwer zu ermitteln
Preisbildung bekannt	Preisbildung nahezu unbekannt
Individueller Besitz; Identifikation und Schutz leicht herstellbar	Vielfacher Besitz möglich; Identifikation und Schutz aufwendig
Hohe Vervielfältigungskosten	Geringe Vervielfältigungskosten
Gebrauch verursacht Wertverlust	Gebrauch generiert Wertgewinn durch Teilung

Tab. 1–2 Gegenüberstellung von materiellen Gütern und Daten als Wirtschaftsgüter

Bekannte Methoden und Technologien zum Asset Management können dementsprechend nicht vollumfänglich – insbesondere im Hinblick auf die Ermittlung des Wertes von Daten – herangezogen werden. Der folgende Abschnitt widmet sich unter der Begrifflichkeit Data Valuation speziell dem Thema Wertermittlung von Daten.

1.3.2 Data Valuation

Die Standards zur Verwaltung von physischen Assets sind in ISO 55001 geregelt. Basierend auf diesem Standard entwickelte The Institute of Asset Management

(IAM) einen Leitfaden zur Bewertung von Wirtschaftsgütern, der als Voraussetzung ein Verständnis über Kosten und Risiken zu einem Vermögensgegenstand über den gesamten Lebenszyklus anführt [Fleckenstein & Fellows 2018, S. 15 ff.].

Als Wirtschaftsgut besitzen Daten einen Wert, der von den Unternehmen zwar erkannt wird, sich allerdings nur schwer bestimmen und quantifizieren lässt. Im International Accounting Standard (IAS) 38 sind immaterielle Wirtschaftsgüter definiert als identifizierbare nicht monetäre Vermögenswerte ohne physische Substanz. Diese Definition trifft auch auf Daten zu, dennoch fließen sie aber bislang nicht als Wirtschaftsgüter in die Bilanzen ein [Treder 2019, S. 43].

Eine Bewertung von Daten kann aus verschiedenen Blickrichtungen und in Abhängigkeit von ihrer spezifischen Rolle in einem Unternehmen durchgeführt werden. Eine gebräuchliche Einteilung unterscheidet zwischen den Bewertungskategorien Kosten, Nutzen und Marktwert. Signifikanten Einfluss auf den Wert von Daten übt die jeweilige Datenqualität aus [Krotova & Spiekermann 2020].

Kosten entstehen entlang des gesamten Data Lifecycle und weisen einen direkten Zusammenhang mit der Bewirtschaftung von Daten auf (z.B. Infrastruktur, Softwarelizenzen, Personal usw.). Eine naheliegende Option besteht darin, die Summe aller angefallenen Kosten für die Erzeugung bzw. Beschaffung und die Pflege der Daten für die Bewertung heranzuziehen. Falls Daten reproduzierbar oder ersetzbar sind, lassen sich alternativ die entsprechenden Reproduktionskosten oder Kosten für einen Datenersatz zugrunde legen [Rea & Sutton 2019, S. 6].

Bezüglich der Kosten seien an dieser Stelle auch Kosten für vermeintlich suboptimale Architekturansätze erwähnt. Lässt eine Architektur z.B. Datensilos zu, dann können Opportunitätskosten entstehen und – im ungünstigsten Fall – sogar die durch Daten gewonnenen Werte zerstören. Als mögliche Folgen von Datensilos lassen sich anführen [Treder 2019, S. 50]:

- Unterschiedliche Antworten auf die gleichen Fragestellungen
- Verzögerung bei der Umsetzung neuer Geschäftsmodelle
- Inkompatibilität bei der Zusammenführung von Silo-übergreifenden Daten
- Doppelte Arbeit
- Erschwerte Einhaltung regulatorischer Anforderungen

Der **Nutzwert** von Daten erweist sich als ungleich schwerer bestimmbar als die zugehörigen Kosten und lässt sich nicht immer exakt quantifizieren. Vor allem wenn neben dem tatsächlich generierten Nutzen (»finanzieller Nutzen«) auch der potenziell mögliche Nutzen (»finanzielle Chance«) erhoben werden soll [Glazer 1993], präsentiert sich die Erhebung als große Herausforderung und lässt sich nur zusammen mit Domänenexperten sowie mit erheblichem Aufwand näherungsweise ermitteln [Krotova & Spiekermann 2020]. Tatsächlich generierter Nutzen erwächst aus dem identifizierten Mehrwert, der sich aus der Datennut-

zung ergibt, beispielsweise durch eine Verbesserung von Prozessabläufen. Daneben gilt es hier, auch Kosten oder entgangene Gewinne zu bestimmen, die aus der Verwendung ungeeigneter oder fehlerhafter Daten resultieren. Potenziell möglicher Nutzen dagegen verweist auf zukünftige Chancen und Risiken durch die Datenverwendung, etwa durch zusätzliche Erlöse oder auch Datenverluste.

Als drittes, mögliches Bewertungskriterium dient der **Marktwert** der Daten, bei dem die Daten als Produkte verstanden und gehandelt werden [Krotova & Spiekermann 2020, S. 30]. Als Voraussetzung hierfür gilt, dass es Marktteilnehmer mit der Bereitschaft gibt, für die Daten zu zahlen. Erst durch das Zusammenreffen von Angebot und Nachfrage für ein Gut erwächst ein Preis. Obwohl bereits erste Datenmarktplätze existieren, die als Plattform einen geregelten Austausch von Daten und dafür zu zahlende Preise gewährleisten wollen, stehen die zugehörigen Geschäftsmodelle noch am Anfang und befinden sich häufig im Aufbau [Krotova & Spiekermann 2020, S. 31]. Den Regelfall dürften daher heute noch bilaterale Austauschbeziehungen zwischen Anbieter und Nachfrager von Daten darstellen. Einen Sonderfall stellt hier der Handel mit Adress- und anderen personenbezogenen Verbraucherdaten dar, der unter engen rechtlichen Rahmenbedingungen gestattet ist [Goldhammer & Wiegand 2017].

Fazit: Dass Daten heute einen ökonomischen Wert besitzen, wird nicht angezweifelt. Auch wenn sich die Bewertungsmethoden noch in der Entwicklungsphase befinden, müssen Daten als Wirtschaftsgut verstanden und entsprechend behandelt werden.

1.3.3 Data Management

Allgemein lässt sich unter Data Management das gesamte Spektrum an technischen, konzeptionellen, methodischen und organisatorischen Methoden, Verfahren und Konzepten zur Steuerung, Beschaffung, Bereitstellung, Verwendung, Qualitätssicherung oder Entsorgung von internen und externen Daten subsumieren. Damit deckt das Data Management den gesamten Lebenszyklus der Daten von ihrer ursprünglichen Erstellung bis zu einem gültigen Ruhezustand ab [Krcmar 2015, S. 178 f.]. Als ausübende Instanz der Datenbewirtschaftung setzt das Data Management die Vorgaben um, die aus diversen Governance-Vorgaben erwachsen [Krotova & Eppelsheimer 2019].

In diesem Kontext erweist sich vor allem das Zusammenspiel von Data Governance und Data Management als richtungsweisend. Obwohl in der betrieblichen Praxis die beiden Begrifflichkeiten häufig als Synonyme Verwendung finden, erweisen sich Data Governance und Data Management als eher komplementär [Al-Ruithe et al. 2018, S. 6]. Als verbindliche Grundlage für die Aktivitäten im Data Management definiert die Data Governance Richtlinien und Prinzipien in den Handlungsfeldern Aufbauorganisation, Prozesse und Standards, Technologie und Kommunikation, die jeweils zu beobachten, zu messen

und zu steuern sind [Gluchowski 2020, S. 6; Fleckenstein & Fellows 2018, S. 63 ff.]. Während die Data Governance damit den Ordnungsrahmen für den angemessenen Umgang mit betrieblichen Daten als wichtige Wirtschaftsgüter aufspannt, setzt das Data Management diese Vorgaben mit geeigneten Konzepten und Werkzeugen um und implementiert sie damit in den Entwicklungs- und Betriebsprozessen [Khatri & Brown 2010].

Hierbei gliedert sich Data Management in eine Reihe von Domänen, die zwar jeweils abgrenzbare Teilaspekte adressieren, allerdings nicht isoliert betrachtet oder gar implementiert werden dürfen, sondern aufgrund vielfältiger Wechselbeziehungen stets im Zusammenspiel mit den anderen Domänen zu betrachten sind [Fleckenstein & Fellows 2018, S. 39]. Beim konkreten Zuschnitt der Domänen und der sich daraus ergebenden Beziehungen entstehen vielfältige Wahlmöglichkeiten, die sich in unterschiedlichen Sichtweisen und Abgrenzungen verfestigen. Den nachfolgenden Ausführungen liegt das »DAMA Wheel« zugrunde (vgl. Abb. 1–4), da es vergleichsweise einfach zu verstehen ist und sich in der Praxis etabliert hat.

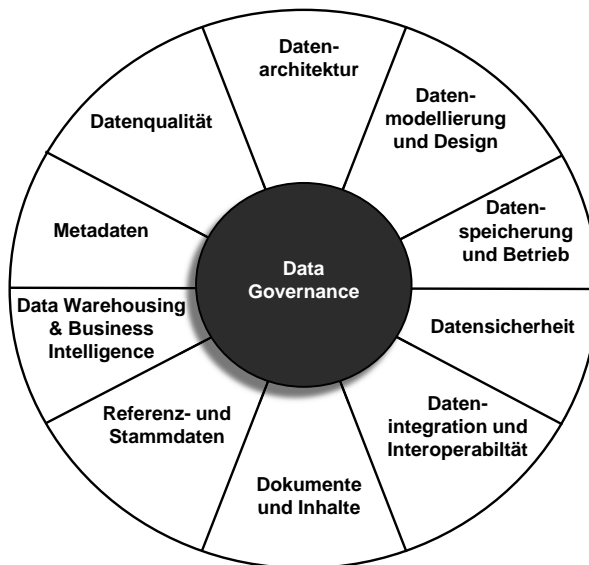


Abb. 1–4 Komponenten des Datenmanagements [DAMA 2017]

Im Ansatz der Data Management Association (DAMA), der unter der Bezeichnung DAMA-Data Management Body of Knowledge (DMBOK) inzwischen in der Version 2 veröffentlicht wurde, setzt sich das Datenmanagement aus elf Komponenten zusammen, von denen die Data Governance den zentralen Ankerbaustein bildet. Der nachfolgende, tabellarische Gesamtüberblick in Tabelle 1–3 leitet über zu einer Beschreibung der für die weiteren Betrachtungen wichtigsten Domänen.

Domäne	Kurzbeschreibung
Data Governance	Orientierung für das Data Management durch Etablierung eines Systems an Verfügungsrechten, Rollen, Prozessen etc.
Data Modeling & Design	Übersetzung von Datenanforderungen in formale Datenmodelle
Data Storage & Operations	Entwurf, Implementierung und Support der Datenspeicherung
Data Security	Management der Vertraulichkeit, Integrität und Verfügbarkeit von Daten
Data Integration & Interoperability	Bewegung und Konsolidierung von Daten zwischen Anwendungen und Organisationen
Document & Content Management	Management des Lebenszyklus von Daten in beliebigen Medien
Reference & Master Data Management	Management der zentralen, geteilten Stammdaten
Data Warehousing & Business Intelligence	Bereitstellung von entscheidungsunterstützenden Daten und Auswertungen
Metadata	Data Management für Metadaten, also beschreibende Angaben über Dateninhalte, -strukturen und -prozesse
Data Quality	Management von Aktivitäten, um Daten in der benötigten Qualität bereitzustellen
Data Architecture	Architektur mit Angaben darüber, wie Daten in ihrem Lebenszyklus durch Systeme fließen

Tab. 1-3 Domänen des Data Management gemäß DAMA Wheel

Aufgabe des **Master Data Management** ist es, dafür Sorge zu tragen, dass die zentralsten Daten einer Organisation wohldefiniert sind [Fleckenstein & Fellows 2018, S. 93]. Hervorgegangen aus dem Management von Kundendaten sowie dem Produktdatenmanagement stehen auch heute diese beiden Entitätstypen im Vordergrund, da sich ihre Bedeutung für die meisten Organisationen als besonders kritisch erweist. Beide Entitätstypen können zwar in komplexen und jeweils hierarchischen Datenmodellen durch verschiedene Sichten unterschiedlicher Fachbereiche münden. Allerdings sind die eigentlichen »Stammdaten« (Master Data) typischerweise klein und beschränkten sich auf wenige Felder mit herausragender und oft übergreifender Bedeutung. Ziel des Master Data Management ist es, für eine konsistente Sicht auf diese Stammdaten in der gesamten Organisation zu sorgen und dabei Dubletten aufzulösen und ggf. die verteilte Bearbeitung der Stammdaten zu ermöglichen. Eng verwandt mit dem Master Data Management ist das **Reference Data Management**, bei dem Daten Beachtung finden, die zum Klassifizieren und Kategorisieren anderer Daten dienen.

In die Domäne **Data Quality** fällt das Management der Datenqualität, das das Ziel verfolgt, Daten in derjenigen Qualität bereitzustellen [Fleckenstein & Fellows 2018, S. 101 ff.], die für die spätere Nutzung erforderlich ist. Der Begriff der Datenqualität kann nicht absolut definiert werden, sondern immer nur in

Abhängigkeit der späteren Anwendung. Typischerweise lassen sich Richtigkeit, Vollständigkeit, Konsistenz, Latenz und Angemessenheit der Daten voneinander abgrenzen. Das Management der Datenqualität umfasst den gesamten Lebenszyklus von Daten, beginnend mit ihrem Entstehen bzw. ihrer Einpflege in ein System. Wichtige Werkzeuge sind z.B. das Data Profiling, das Data Quality Monitoring sowie das Data Cleansing, wobei sich in diesen Tätigkeitsfeldern die Einbindung von Mitarbeitern aus den Fachabteilungen als erforderlich erweist. Eine hohe Datenqualität erfordert dabei grundsätzlich die übergreifende Zusammenarbeit, Kommunikation und auch Abstimmung. Maßnahmen zur Qualitätsverbesserung von Daten orientieren sich häufig an Vorgehensweisen zur Qualitätsverbesserung in industriellen Prozessen, folgen damit z.B. dem Deming- oder PDCA-Zyklus (bestehend aus Plan, Do, Check, Act) [Deming 1982].

Die Domäne **Data Security** soll gewährleisten, was häufig mit dem Akronym C.I.A. bezeichnet wird: die Vertraulichkeit (Confidentiality), Integrität (Integrity) und Verfügbarkeit (Availability) von Daten [Fleckenstein & Fellows 2018, S. 166]. Wie in anderen Managementdisziplinen auch, besteht die Aufgabe darin, den gewünschten Status zu planen (z.B. Kategorisierung von Daten nach Schutzbedürftigkeit und Festlegung dessen), diesen umzusetzen, die Einhaltung zu überwachen bzw. zu auditieren und eventuelle Auffälligkeiten in geeigneter Weise zu behandeln.

Metadata umfasst beschreibende Angaben über den Inhalt, die Struktur, die Verarbeitung und die Nutzung von Daten und bildet damit eine Grundvoraussetzung für die effektive Verwendung des Datenbestands einer Organisation [Fleckenstein & Fellows 2018, S. 166]. So wie »Data Management« den professionellen Umgang mit Daten allgemein adressiert, bezieht sich »Metadata Management« auf die Teilmenge der Metadaten. Metadaten vereinfachen die Zugänglichkeit von Daten, z.B. durch die Verwendung eines Datenkatalogs, der Daten-Assets in einem zentralen Verzeichnis verwaltet. Sie helfen mit Einblicken in die Datenherkunft (Data Lineage: Rückverfolgung der Herkunft von Daten von der Auswertung zur Quelle) und Datenverwendung (Impact Analysis: Abschätzung von Auswirkungen in nachgelagerten Systemen bei Veränderungen in Quellsystemen). Darüber hinaus lassen sich Metadaten effizienzsteigernd bei der Automatisierung verwenden (bspw. im Rahmen des Data-Vault-Ansatzes).

Ziel der Domäne **Data Modeling & Design** ist es, die genauen Datenanforderungen zu verstehen und diese anschließend in ein formales Datenmodell – besteht aus Metadaten zur Beschreibung der konkreten Daten – zu überführen [DAMA 2017, Kap. 5]. Je nach prinzipieller Struktur der Daten gibt es eine Reihe unterschiedlicher Schemata und Notationen für Datenmodelle, darunter UML für objektorientierte Strukturen, Entity-Relationship-Diagramme z.B. in der Chen-Notation für relationale Strukturen oder die Data-Vault-Modellierungstechnik.

Mit **Data Integration & Interoperability** werden Prozesse beschrieben, im Rahmen derer Daten über System-, Anwendungs- und/oder Organisationsgrenzen hinweg transportiert oder transformiert werden [DAMA 2017, Kap. 8]. Zu

den wesentlichen Konzepten zählen das Batch-orientierte ETL bzw. ELT genauso wie Streaming-Ansätze, ein Enterprise Service Bus (ESB) und Data Virtualization.

Mit den einzelnen Facetten des DAMA Wheel sind auch die Handlungsfelder für das Datenmanagement in Unternehmen umrissen, die allerdings auch mit den Zielvorstellungen und Prioritäten der unterschiedlichen Stakeholder in Deckung gebracht werden müssen, wie der folgende Abschnitt aufzeigt.

1.4 Anforderungen an eine ganzheitliche BIA-Architektur

Die Gestaltung einer angemessenen Architektur für BIA-Landschaften muss sich an den gegebenen Rahmenbedingungen orientieren. Dazu gehören neben der grundsätzlichen Ausrichtung der jeweiligen Organisation mit Geschäftsmodell und Unternehmensstrategie auch die darauf aufbauenden Vorgaben, die in der Datenstrategie und der Data Governance dokumentiert sind.

Als weiterer wichtiger Bestimmungsfaktor sind die funktionalen und nicht funktionalen Anforderungen der einzelnen Anspruchsgruppen anzuführen. Funktionale Anforderungen werden in erheblicher Weise vom jeweiligen Anwendungskontext determiniert und variieren stark in Abhängigkeit von Branchenzugehörigkeit, Unternehmensgröße und zu unterstützender Unternehmensfunktion. Dagegen adressieren die nicht funktionalen Anforderungen – auch als Qualitätsanforderungen bezeichnet – Themengebiete wie Performance, Verfügbarkeit, Zuverlässigkeit, Skalierbarkeit und Portabilität [Pohl & Rupp 2011, S. 16].

Als Anspruchsgruppen lassen sich neben dem Management grob die Anwender von den Entwicklern und Betreibern unterscheiden (vgl. Abb. 1–5). Das Management setzt sich in dieser Einteilung aus den Entscheidungsträgern oberer und ggf. mittlerer Führungsebenen zusammen und nimmt einen eher übergeordneten Blick auf die Systemlandschaft ein (auch wenn Führungskräfte als Anwender auf bestimmte Teile des BIA-Ökosystems zugreifen wollen). Den Entscheidern ist insbesondere die Einhaltung übergeordneter interner und externer Regeln und Vorgaben wichtig, weshalb hier die Themenbereiche Compliance und Governance stark im Fokus stehen. Als Budgetverantwortliche stecken sie den Rahmen für die verfügbaren finanziellen Ressourcen ab und richten besonderes Augenmerk auf Nutzen und Kosten (Wirtschaftlichkeit) eines BIA-Vorhabens.

Den Anwendern von BIA-Lösungen, die zumeist in den einschlägigen Fachbereichen wie Controlling oder Marketing beheimatet sind, liegt vor allem die bestmögliche Unterstützung der eigenen, fachlichen Arbeitsaufgaben am Herzen. Immer stärker rückt dabei die Geschwindigkeit in den Vordergrund, mit der sich neue fachliche Anforderungen im geschäftlichen Kontext ergeben und die rasch in den Systemen ihren Niederschlag finden sollen. Da sich eine derartig ausgeprägte Agilität nicht immer mit den Verfahren eines klassischen Anforderungs- und Projektmanagements erreichen lässt, wird häufig größere Autonomie bei der selbstständigen Erarbeitung von Lösungen eingefordert, die in Konzepten wie

Self-Service-BI ihre Umsetzung findet. Um dem schnelllebigen fachlichen Umfeld gerecht werden zu können, bedarf es auch Lösungen, die mit möglichst geringem Zeitverzug (Latenz) neue Informationen zugänglich machen. Zur Bewältigung dieser Herausforderung lassen sich Real- bzw. Right-Time-Konzepte verwenden. Daneben werden hier Werkzeuge zur Automatisierung eingesetzt, die zusätzlich von aufwendigen Vorarbeiten zur Datenbereinigung und -transformation entlasten können. In Bezug auf die auswertbaren Datenformate hat seit einigen Jahren eine verstärkte Einbeziehung auch unstrukturierter sowie semistrukturierter Daten in den Fachbereichen Einzug gehalten – teils aus internen (Sensor Daten), teils aus externen Quellen (Daten aus sozialen Medien). Unabhängig von der Datenherkunft erwarten die Anwender oftmals, dass sie den angebotenen Inhalten vertrauen können.

In Abgrenzung vom Anwender verfolgen die Entwickler und Betreiber von BIA-Landschaften oftmals abweichende Ziele, aus denen eine andere Priorisierung von Anforderungen resultieren kann. Häufig erhalten hier beispielsweise Aspekte wie Datensicherheit und -schutz eine viel höhere Bedeutung wie auch die Widerspruchsfreiheit der Daten (Konsistenz) und umfangreiche Transparenz über die einzelnen Objekte der Architektur einschließlich Strukturen und Abhängigkeiten (z.B. Datenflüsse). Als wichtige Faktoren gelten weiterhin die Skalierbarkeit der Systemlösung, beispielsweise bei steigendem Datenvolumen, sowie die jederzeitige Verfügbarkeit und Stabilität.

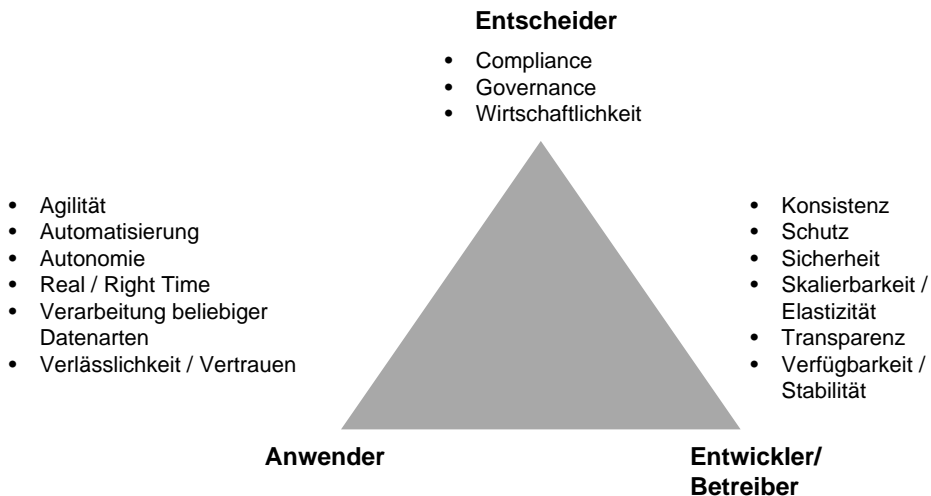


Abb. 1-5 Anforderung von Anspruchsgruppen an eine BIA-Landschaft

Die Ausführungen zeigen, dass die unterschiedlichen Anspruchsgruppen stark voneinander abweichende Zielvorstellungen und damit auch Anforderungen an eine geeignete BIA-Systemlandschaft aufweisen.

Der folgende Abschnitt beleuchtet nochmals die klassischen BIA-Architekturen, die über viele Jahre hinweg in den Organisationen als stabiles Fundament für die Gesamtlanschaft genutzt worden sind und bisweilen auch heute noch genutzt werden.

1.5 Klassische Architekturen für BIA-Ökosysteme

Bevor der Hauptteil des vorliegenden Sammelbandes auf moderne BIA-Architekturen eingeht, soll an dieser Stelle zunächst ein Blick zurück geworfen werden auf die Architekturformen, die sich langjährig als Standard im Bereich BIA etablieren konnten.

Für eine gewisse Zeit schien die Diskussion um BIA-Architekturen bis auf Nuancen erledigt zu sein. Nach anfänglich konträren Ansätzen hatte sich schließlich die Hub-and-Spoke-Architektur als Quasistandard etabliert [Hahne 2014, S. 10 ff.]. Zugrunde liegen dabei die Kernideen eines Data Warehouse (DWH) durch Trennung von operativen und dispositiven Daten und Zusammenführung von Daten aus unterschiedlichen Quellen in einem abgestimmten Datenmodell (vgl. Abb. 1–6).

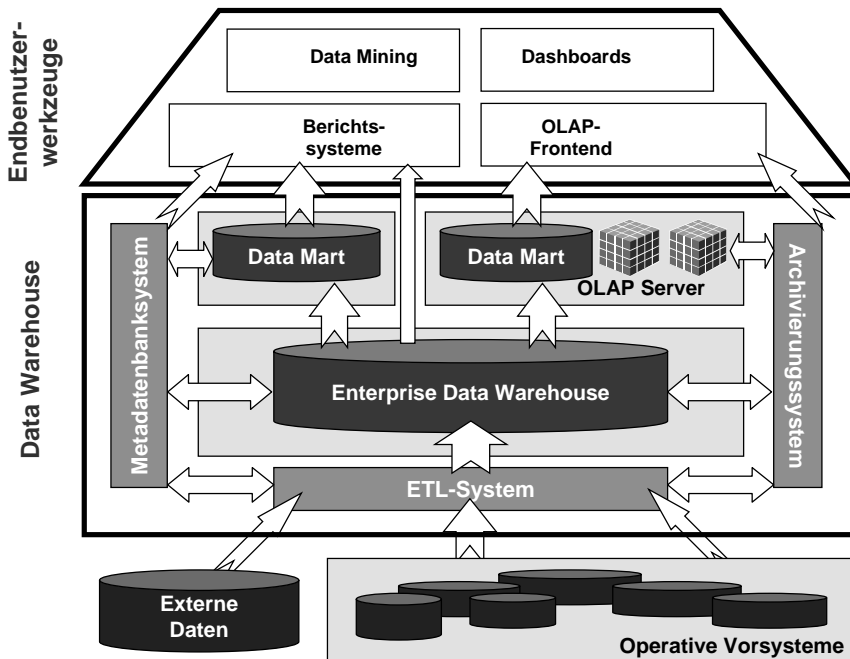


Abb. 1–6 Hub-and-Spoke-Architektur [Gluchowski et al. 2008, S. 141]

Das Data Warehouse hält Schnappschüsse des unternehmensweiten Datenstands von operationalen Systemen dauerhaft fest und stellt diese für Analysen bereit [Fleckenstein & Fellows 2018, S. 123]. Das DWH-Konzept kam in den 1990er-Jahren erstmalig zur praktischen Anwendung und dient seither als Datenbasis für strategische Berichts-, Analyse- und Planungssysteme [Bauer & Günzel 2013, S. 11–13].

Neben kostenlosen oder kostenpflichtigen externen Daten dienen als primäre Datenquellen die operativen Systeme. Per ETL (Extraktion, Transformation, Laden) werden hieraus Daten über potenziell mehrere Zwischenschichten in ein passendes Schema überführt und im Enterprise Data Warehouse gespeichert. Neben dieser Vorgehensweise bietet sich insbesondere in großen Landschaften mit hohem Datenvolumen ein ELT an, bei dem nach der Extraktion die Daten zunächst in einen speziellen Speicherbereich des Data Warehouse (Staging Area) geschrieben und danach meist mit datenbankeigenen Mitteln wie Views oder Skripten weiterverarbeitet werden.

Anschließend erfolgt die Bereitstellung von Teilmengen der Daten aus dem Data Warehouse für die Fachanwender als Data Marts, dann z.B. zugeschnitten auf die Anforderungen eines spezifischen Fachbereichs. Diese Data Marts dienen schließlich als Datenquelle für die Auswertungen mit geeigneten Endbenutzerwerkzeugen, z.B. einem Berichtssystem oder Softwareprodukten für die Visualisierung der Daten als Dashboard.

Die Bezeichnung Hub-and-Spoke (Nabe und Speiche) wird allgemein für Architekturen verwendet, wenn die Verbindung von der Quelle zur Senke nicht direkt, sondern über einen zentralen Verbindungsknoten führt, die Nabe (engl. Hub). Im konkreten Fall der klassischen BIA-Architektur werden die Daten ausgehend von den Quellsystemen über die zentrale Nabe Data Warehouse in die Data Marts übertragen, von wo aus sie die unterschiedlichen Auswertungswerkzeuge speisen.

Eine Metadatenverwaltungskomponente sowie eine Archivierungs- und Backup-Komponente flankieren die Kernkomponenten der Hub-and-Spoke-Architektur.

Ein näherer Blick auf das Data Warehouse selbst enthüllt verschiedene logische (Daten-)Schichten bzw. Layer (vgl. Abb. 1–7).

Wie bereits vorangehend illustriert, integriert das Enterprise Data Warehouse Inhalte aus (i.d.R. mehreren) Datenquellen und leitet sie schlussendlich an (i.d.R. mehrere) Data Marts bzw. den Reporting Layer für Auswertungen weiter.

Eine zentrale Sammlung der Daten erfolgt im Integration Layer oder Core bzw. Enterprise Data Warehouse. Meist werden die Daten hier heute in größtmöglicher Detaillierung gespeichert und über mehrjährige Zeiträume aufbewahrt. Im Idealfall existiert für die gesamte Organisation nur ein einzelner Integration Layer – dann auch als »Single Point of Truth« (SPOT) bezeichnet – mit aufbereiteten und qualitätsgesicherten Inhalten.

Zwischen den Datenquellen und dem Integration Layer befindet sich der Acquisition Layer, häufig auch als Staging Area bezeichnet. Dieser Layer wird genutzt, um Daten zunächst unverändert aus den Quellsystemen zu kopieren. Nachfolgend lassen sich dann Transformationen zwischen dem Acquisition Layer und dem Integration Layer realisieren (ELT), um die Daten in die gewünschte Struktur zu bringen. Diese Transformationen erfolgen dann innerhalb des Data Warehouse und belasten daher nicht die Quellsysteme.

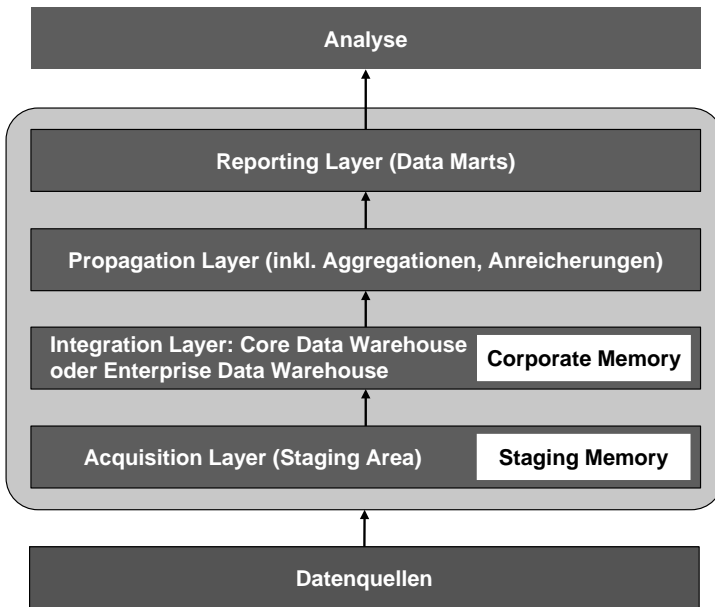


Abb. 1-7 Data-Warehouse-Schichtenarchitektur

Der Propagation Layer setzt auf den Integration Layer auf und befindet sich vor dem Reporting Layer. Kerngedanke dieser Schicht ist, dass gewisse Geschäftslogiken, Aggregationen und Anreicherungen für unterschiedliche Data Marts in identischer Form benötigt und demzufolge nur einmal realisiert werden sollten. Die Bündelung dieser Logiken hilft der Qualitätssicherung zur Sicherung der Konsistenz der Logiken und reduziert gleichzeitig den Aufwand durch die Wiederverwendbarkeit.

Zwar hat das klassische Hub-and-Spoke-Konzept mit der zugehörigen Schichtenarchitektur über lange Jahre und mit ausgeprägter Stabilität gute Dienste geleistet, allerdings sind neue Anforderungen erwachsen, die eine Veränderung oder zumindest Erweiterung dieses Aufbaus erforderlich machen. So sind die klassischen DWH-Architekturen mit den zugrunde liegenden relationalen Datenbanksystemen auf den Umgang mit strukturierten Daten in Tabellenform ausgerichtet. Zusätzliche Datenformate zur Speicherung von Audio- oder Videodaten wie auch unstrukturierte Textdaten lassen sich dagegen nur sehr ineffizient verar-

beiten. Weiterhin ist die Architektur auf die Integration neuer Inhalte im Batch-Betrieb ausgelegt, meist mit nächtlichen Aktualisierungsprozessen. Eine untertägige Ergänzung der Datenbestände im zentralen Data Warehouse – vielleicht gar in Echtzeit – erfordert jedoch einen abweichenden Aufbau und zusätzliche Technologien. Zudem sind die zu verarbeitenden Datenmengen in den letzten Jahren kontinuierlich angewachsen, was es erforderlich macht, statt auf eine zentrale eher auf eine verteilte Datenhaltung zu setzen. Darüber hinaus haben sich neue Technologien etabliert, die eine kostengünstigere oder leistungsfähigere Verarbeitung des Datenmaterials eröffnen.

Obleich die klassischen Architekturen für zahlreiche Jahre den Standard gebildet haben, müssen moderne BIA-Architekturen den neuen Herausforderungen Rechnung tragen, die zuvor geschildert wurden. Die folgenden Kapitel nähern sich – ausgehend von einigen grundlegenden und herstellerneutralen Betrachtungen – modernen BIA-Architekturen weiter an, indem idealtypische Gestaltungsformen unterschiedlicher Produkthanbieter aufgezeigt und diskutiert werden. Anschließend illustrieren Beispiele konkrete Anwendungsfälle. Ein wirklicher Standard hat sich hier noch nicht herauskristallisiert, sodass die Beispiele nur als Momentaufnahmen einer laufenden Evolution und Entwicklung zu verstehen sind.

2 Architekturen und Technologien für Data Lakes

Carsten Dittmar • Peter Schulz

Lange Zeit galt das Data Warehouse als zentrales Architekturkonzept für dispositive Reporting- & Analyse Zwecke. Im Zuge der zunehmenden Digitalisierung und der damit einhergehenden Masse an zur Verfügung stehenden Datenmengen, aber auch des breiten Spektrums an potenziellen datenbasierten Use Cases hat mittlerweile der Data Lake dem klassischen Data Warehouse den Rang abgelaufen. In diesem Beitrag wird nach einer Einordnung das Konzept des Data Lake vorgestellt und anschließend werden gängige Architektur- und Technologiemonster der Praxis skizziert.

2.1 Historie der dispositiven Datenplattformen

Die richtige Architektur für die dispositive Datenverarbeitung war seit den 90er-Jahren mit dem Data Warehouse klar definiert. Mit dem Data-Warehouse-Konzept setzte sich die Idee einer separaten Datenbasis für dispositive Reporting- und Analyse Zwecke in der Praxis durch, die redundant Datenbestände aus den operativen Systemen speichert. Bisherige Architekturansätze, die zu Reporting- und Analyse Zwecken direkt auf die operativen Systeme und ihre Daten zugriffen, waren mit dem Data Warehouse obsolet. Die potenziellen Gefahren einer Inkonsistenz der redundanten dispositiven Datenbestände gegenüber den operativen Datenbeständen wurden durch einen gesteuerten Datenintegrationsprozess in das Data Warehouse sowie einer Beschränkung auf einen Lesezugriff auf das Data Warehouse beantwortet [Gluchowski et al. 2008, S. 128 ff.].

Die Standardarchitektur sieht idealtypisch ein singuläres (Enterprise) Data Warehouse vor, das aus den unterschiedlichen operativen Quellsystemen die relevanten Daten aufammelt. In einem mehrschichtigen Datenintegrations- und Datenveredelungsprozess werden die Daten im Data Warehouse harmonisiert, integriert und persistiert. So soll aus einer Datenperspektive ein Single Point of Truth entstehen, aus dem anschließend für unterschiedliche Anwendungsfälle Datenextrakte – in der Regel in einer multidimensionalen Aufbereitung – in abgrenzbaren Data Marts gespeichert werden [Schnider et al. 2016, S. 6 ff.].

Zwei architekturelle Grundmuster haben sich zur Umsetzung der mehrschichtigen Data-Warehouse-Architektur bewährt. Der Hub-and-Spoke-Ansatz nach Inmon gilt nach wie vor als die Reinform des Data Warehouse, da in der Schicht des Core Data Warehouse eine integrierte und umfassende (Unternehmens-)Datensicht erstellt wird [Inmon 2005]. Als pragmatische Architekturalternative schreibt man gemeinhin dem Data Mart Bus nach Kimball Vorteile bei der schnellen Umsetzung und Erweiterbarkeit zu, da der Fokus einer Harmonisierung nur auf die strukturgebenden Stammdaten gelegt wird [Kimball & Ross 2002].

Auf den Datenschatz des Data Warehouse greift der Anwender selten direkt zu. Zumeist nutzt er Berichts- und Analysewerkzeuge (Business Intelligence), die den Zugriff auf Standardberichte erlauben, aber mittlerweile dem User auch die Möglichkeiten geben, im Self-Service eigene Datenanalysen auf Basis individueller Datenzusammenstellungen durchzuführen. Der überwiegende Fokus liegt im Data Warehouse auf der vergangenheitsorientierten Analyse von Kennzahlen entlang von konsolidierten Auswertungsstrukturen. Damit werden in der Regel Fragestellungen wie »Was ist passiert?« und »Warum ist es passiert?« adressiert.

Um eine gültige und konsolidierte Wahrheit zu allen strukturierten Daten in einem Unternehmen zu repräsentieren, stellt das Data Warehouse die Daten in vorab definierten Datenmodellen zur Verfügung. Bevor Daten integriert werden, ist demzufolge dieses Datenmodell zu entwickeln und zu implementieren. Der hohe Anspruch an korrekte und unternehmensweit harmonisierte Daten führt in der Regel dazu, dass es recht lange dauert, bis Daten aus einer neuen Datenquelle in dieser konsolidierten Sicht integriert sind, weil im Vorfeld viel Konzeptions- und Abstimmungsaufwand nötig wird.

Bestrebungen, eine dispositive Datensenke neben den operativen Systemen obsolet werden zu lassen und alle operativen und auch dispositiven Anfragen mit demselben System beantworten zu können, hat insbesondere das SanssouciDB-Vorhaben des Hasso-Plattner-Instituts und der Stanford University vorangetrieben [Plattner & Zeier 2012]. Das daraus resultierende kommerzielle Produkt ist die SAP HANA-Datenbank, die in der aktuellen Version der Enterprise-Resource-Planning-(ERP-)Lösung des Herstellers eingesetzt wird. Diese Lösung stellt für einige Data-Warehouse-Fragestellungen eine tatsächliche Alternative dar, kann Daten außerhalb der operativen Applikation aber nur schwer integrieren.

2.2 Das Data-Lake-Konzept

Im neuen Jahrtausend stieg mit dem Aufkommen neuer Datenquellen wie Social-Media-Daten oder IoT-Daten und dem enormen Anstieg des Datenvolumens durch die Digitalisierung vieler Prozesse auch der Bedarf, diese neuen Daten ebenfalls in einer zusammenführenden Datenplattform zur Verfügung zu stellen. Viele dieser Daten liegen jedoch in semistrukturierter oder unstrukturierter Form vor. Mit der steigenden Relevanz dieser Datenquellen wurde die Idee des Data

Lake geboren. Die Idee wird gemeinhin James Dixon zugeschrieben, der in einem Blogpost von 2010 erstmalig das Bild eines Data Lake prägt [Dixon 2010]. Der Data Lake stellt alle Quelldaten – interne und externe, strukturierte und unstrukturierte – auch in ihrer nicht aufbereiteten Form als Rohdaten zur Verfügung. Somit stehen die Daten möglichst unmittelbar nach der Datenerzeugung schnell und unverfälscht in einem Data Lake bereit. Dadurch werden Einblicke zur Echtzeit ermöglicht, die auf Basis von Vorhersage- und Szenariomodellen die Fragestellungen »Was wird wahrscheinlich passieren?« und »Was kann unternommen werden, damit es passieren wird?« beantworten können.

Die Speicherung von Rohdaten ohne jegliche Datenveredelung auf feinsten Granularität oblag in einer klassischen Data-Warehouse-Architektur eher der Datenschicht der Staging Area. Daher wird häufig der Data Lake als Fortentwicklung dieser Schicht gesehen.

Der effiziente Umgang mit großen polystrukturierten Datenmengen, eine schnelle (oft nahezu in Echtzeit) Verarbeitung von Datenströmen und die Beherrschung komplexer Analysen für neue Data-Science- und Machine-Learning-Anwendungen stehen beim Data Lake zulasten der Harmonisierung und Integration der Daten im Vordergrund. Die Struktur der Daten steht damit zugunsten einer schnellen und vollständigen Integration in den Data Lake nicht schon bei der Speicherung, sondern erst im Rahmen der nachgelagerten Analyse im Fokus. Somit ist das Ziel eines Data Lake die Schaffung von flexiblen Strukturen zur Bändigung der komplexen Integration der Vielzahl von Datenquellen.

Zumeist ist zum Zeitpunkt der Datenspeicherung noch gar nicht festgelegt, welche Analysen mit den Daten durchgeführt werden sollen. Der Data Lake bildet also das Eldorado für den Data Scientist, der explorative Analysen wie Cluster-/Assoziationsanalysen, Simulationen und Vorhersagen über komplexe Algorithmen durchführen möchte. In der folgenden Tabelle 2–1 sind wesentliche Charakteristika des Data Warehouse und des Data Lake vergleichend zusammengefasst.

Data Warehouse: Datenbasis für Systems of Record	Data Lake: Datenbasis für Systems of Innovation
<ul style="list-style-type: none"> ■ Stellt 80 % der Analysen mit 20 % der Daten bereit ■ Optimiert für wiederholbare Prozesse ■ Unterstützt Vielzahl von unternehmens-internen Informationsbedarfen ■ Fokus auf vergangenheitsorientierte Auswertungen ■ Schema-on-Write mit harmonisiertem Datenmodell 	<ul style="list-style-type: none"> ■ Originäre Erweiterung der Staging Area des DWH ■ Speichert Rohdaten für Exploration und Analyse ■ Optimiert Daten unkompliziert für Analytics-Lösungen ■ Fokus auf unbekanntes Data Discovery und zukunftsorientierte Data Science & Artificial Intelligence ■ Schema-on-Read mit Echtzeit-Rohdatenbewirtschaftung

Tab. 2–1 Charakteristika von Data Warehouse und Data Lake im Vergleich

Damit ist auch klar, dass ein Data Lake ein Data Warehouse nicht ersetzt, sondern ergänzt. Beide Architekturkonzepte haben ihre Relevanz und bedienen zueinander unterschiedliche Use Cases [Gartner 2020].

Verschiedene übergreifende Gesamtarchitekturen sind denkbar. Sofern der Data Lake als übergreifender Staging Layer fungiert, folgt in einer sequenziellen Architekturabfolge das Data Warehouse hinter dem Data Lake. In der Regel stehen jedoch beide Systeme parallel und isoliert nebeneinander. Häufig wird im letzten Falle über Virtualisierungstechnologie für den Anwender ein virtueller Datenmarktplatz geschaffen, der beide Architekturkonzepte vermeintlich vereint [Leisten 2020].

Aktuell Verbreitung findet die Variante des Data-Mesh-Konzeptes, das neben seinen Architekturaspekten auch Governance und prozessuale Aspekte abdeckt [Dehghani 2020].

Der Reifegrad der Nutzung des Data-Lake-Konzeptes kann anhand der Dimensionen des verwendeten Datenhaushaltes und der Nutzung unterschieden werden [Gorelik 2019, S. 9 ff.]. Dieses Modell nimmt Bezug auf die ursprünglich von James Dixon aufgestellte Definition: »Wenn Sie sich Data Mart als einen Vorrat an Wasser in Flaschen vorstellen – gereinigt, verpackt und strukturiert für einen einfachen Verbrauch, ist der Data Lake ein großes Gewässer in einem natürlicheren Zustand. Der Inhalt des Datensees fließt von einer Quelle herein, um den See zu füllen, und verschiedene Benutzer des Sees können kommen, um zu untersuchen, einzutauchen oder Proben zu entnehmen« ([Dixon 2010], eigene Übersetzung aus dem Englischen). Die nachfolgende Abbildung 2–1 verdeutlicht die fünf unterschiedlichen Ausprägungen.

Ein Data Puddle ist die erste Adaptionsform des Data-Lake-Konzeptes. Er ist gekennzeichnet durch seinen eingeschränkten, nur den aktuellen Nutzungsfall abdeckenden Datenhaushalt und eine lokale Nutzung durch (IT-)Experten, die einen hohen Grad an manuellen Tätigkeiten zur Nutzung bedingt. Hier stehen Kostensenkung und höhere Performance im Vergleich zum Einsatz klassischer Technologien eines Data Warehouse im Vordergrund, ein Mehrwert aus der Nutzung gegenüber einem Data Warehouse wird jedoch nicht geschaffen.

Als Data-Pond-Konstrukt wird eine Vielzahl von nebeneinander isoliert bestehenden Data Puddles bezeichnet. Ein gängiges Beispiel ist die Kopie von mehreren Datenhaushalten aus Data Warehouses in separate Systeme, die auf typischen Data-Lake-Technologien basieren. Neue Erkenntnisse sind auch mit diesem Archetyp nur sehr eingeschränkt und umständlich zu gewinnen.

Ein Data Lake unterscheidet sich von einem Data Pond in zwei wesentlichen Faktoren: Erstens ermöglicht er Self-Service-Nutzung durch Anwender ohne IT-Beteiligung und zweitens enthält er Daten, die aktuell noch gar nicht genutzt werden, aber perspektivisch interessant werden können. Die breite Nutzung im Unternehmen wird maßgeblich dadurch erreicht, dass Daten für die Nutzung aufbereitet und vor allem durch (einfach zugängliche) Metadaten beschrieben sind. Mit dieser Aufstellung können von einer Vielzahl unterschiedlicher Nutzergruppen Antworten auf ihre individuellen Fragestellungen gewonnen werden.

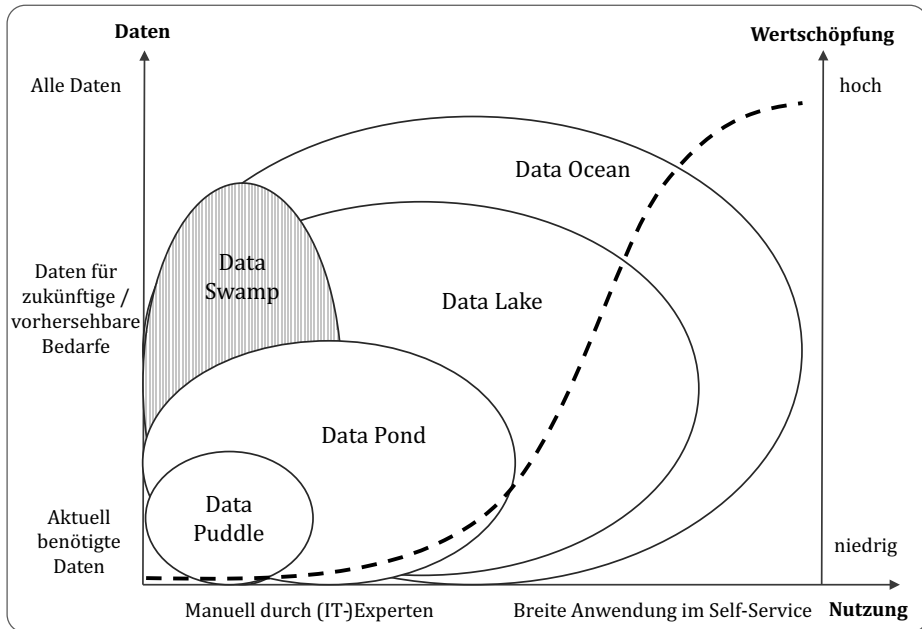


Abb. 2-1 Reifegrad der Nutzung des Data-Lake-Konzeptes und resultierende Wertschöpfung (gestrichelte Linie) (eigene Darstellung in Anlehnung an [Gorelik 2019])

Ein Data Ocean gilt als ultimative Antwort auf datengetriebene Entscheidungen, basierend auf allen (fachlichen) Daten eines Unternehmens und mit einem einfachen, verständlichen Zugang für alle Mitarbeiter. Die resultierende Wertschöpfung kann jedoch gegenüber einem gut positionierten Data Lake nur noch marginal erhöht werden.

Eine Sonderform stellt der berüchtigte Data Swamp dar. Er ist eine Ansammlung von verschiedenen Daten, die jedoch überhaupt nicht oder wenig organisiert und aufbereitet sind. Weiterhin fehlen Metadaten, was eine Nutzung durch eine breitere Anwenderbasis verhindert. In der Praxis wird in einem solchen Fall die Behandlung mit dem Ziel der Umwandlung in einen Data Lake durch Maßnahmen wie Aufbereitung der Daten und Zuordnung von Metadaten versucht.

2.3 Architektur eines Data Lake

Die idealtypische Architektur eines Data Lake durchlief eine Evolution: Glaubte man zunächst, eine einzige, große Plattform für alle Bedürfnisse im eigenen Rechenzentrum zu bauen sei optimal, hat sich hier ein Trend zur Nutzung multipler, aber orchestrierter Data Lakes und auch die Nutzung von Cloud-Angeboten durchgesetzt.

Ein Katalog an Architekturprinzipien für einen Data Lake ist in der nachfolgenden Tabelle 2-2 dargestellt.

14 BIA-Architekturen für kleine und mittlere Unternehmen

Markus Begerow

Dieser Beitrag setzt sich zum Ziel, die bevorstehenden Herausforderungen für kleine und mittlere Unternehmen (KMU) im digitalen Zeitalter herauszuarbeiten. Den Ausgangspunkt für die Betrachtungen bildet der folgende erste Abschnitt, der einen Rückblick und den Status quo zur aktuellen Ausgangssituation in den letzten Jahren für KMU im Allgemeinen beschreibt. Danach erfolgt eine Abgrenzung und Einordnung der aktuellen Ausgangssituation von KMU in Abhängigkeit zu neuen Themen, die als Treiber für Innovationen stehen (Abschnitt 14.2). Der anschließende Abschnitt 14.3 beleuchtet die Konsequenzen für KMU, die mit den neuen Innovationen einhergehen. Dabei wird nach möglichen Szenarien und Vorgehensweisen zur Umsetzung, inklusive der Abwägung von Für und Wider, unterschieden. Abschließend greift Abschnitt 14.4 die zuvor genannten Argumente auf und bewertet diese in einem Fazit mit Ausblick auf die Herausforderungen.

14.1 Ausgangssituation

In Zeiten von Globalisierung, Pandemien, wachsenden Märkten, immer schnelleren Marktteilnehmern und wechselnden technologischen Trends steigen auch die strategischen und operativen Anforderungen an Unternehmen.

Ein Unternehmen, das sich weiterhin die Wettbewerbsvorteile sichern möchte, muss auf Marktveränderungen schnell reagieren sowie Chancen erkennen, Risiken und Kosten minimieren und entsprechend handeln. Die Grundlage für zielsichere Unternehmensentscheidungen bilden häufig aussagekräftige Informationen, die aus den im Unternehmen vorhandenen Daten abgeleitet werden.

Der rapide Anstieg der Datenmengen, besonders in den letzten Jahren, hat in Unternehmen jedoch dazu geführt, dass die manuelle Aufbereitung und Harmonisierung der Informationen für eine sinnvolle, übergreifende Informationsentscheidung kaum noch möglich sind. Hier können Business-Intelligence-Systeme unterstützend Abhilfe schaffen. Mit bestimmten Werkzeugen und entsprechenden Technologien werden die Fachbereiche bei der Sammlung, Aufbereitung und Analyse der vorhandenen Daten unterstützt. Während das Thema in den letzten

Jahren hauptsächlich in Großunternehmen interessant war und diskutiert wurde, ist es inzwischen auch für die KMU sehr wichtig geworden.

Viele KMU stehen heute aber auch anderen Herausforderungen gegenüber, die vor einigen Jahren noch eher als uninteressant abgehakt wurden. Die Themen Planung und Konsolidierung beispielsweise gewinnen eine immer größere Bedeutung. Als Konsequenz aus der letzten Finanzkrise unterliegen Banken heute strengerem Kreditvergaberichtlinien und legen mehr Wert auf nachprüfbar Größen wie Bilanzkennzahlen oder ein funktionierendes Berichtswesen im Controlling.

Eine weitere Herausforderung für die KMU ist, auch weiterhin in den bereits erwähnten Märkten zu bestehen und sich dort behaupten zu können. Damit zukünftig Marktchancen abgeschätzt werden können, müssen Markt-, Kunden- und Wettbewerberverhalten analysiert, ausgewertet und in Relation zum eigenen Unternehmen gestellt und die Ergebnisse interpretiert werden.

Auch die Wettbewerbssituation hat sich für die KMU durch die Globalisierung stark verändert. Besonders regionale Unternehmen stehen heutzutage immer häufiger in direkter Konkurrenz zu global agierenden Konzernen.

Ein funktionierendes Berichtswesen kann in KMU bei der Bewältigung dieser Herausforderungen helfen und eine Entscheidung erheblich beschleunigen. Aussagekräftige Berichte und gute Auswertungen, die auf verlässlichen und aktuellen Daten basieren, können als Grundlage für Entscheidungen dienen und die Reaktionsfähigkeit, Effektivität und Wettbewerbsfähigkeit der Unternehmen steigern.

Neben den bereits genannten Herausforderungen haben sich nun neue Themen als Treiber für Innovationen herauskristallisiert, die von KMU bewältigt werden müssen, um weiterhin am wirtschaftlichen Geschehen teilzunehmen.

14.2 Neue Themen als Treiber für Innovationen

In diesem Unterkapitel werden die wesentlichen fachlichen und technologischen Treiber für Innovationen vorgestellt. In jedem Abschnitt erfolgt zunächst die Beschreibung der grundlegenden Situation. Anschließend wird die Betrachtungsweise für KMU eingenommen, um daraus mögliche Herausforderungen abzuleiten.

14.2.1 Stammdaten- und Datenqualitätsmanagement

Das Thema *Stammdatenmanagement* ist nicht nur Gegenstand einiger Großunternehmen, sondern begegnet auch immer mehr KMU in ihren täglichen Routinen.

Stammdaten sind viel langlebiger als Bewegungsdaten. Daher erfordert ihre Verwaltung eine kontinuierliche Steuerung, Wartung und Pflege, und das über den gesamten Lebenszyklus hinweg, von der Anlage bis zur Löschung.

Die Harmonisierung dieser Stammdaten und die Synchronisierung der unterschiedlichsten Datenquellen ist eine komplexe Herausforderung, die von KMU nur schwer zu meistern sein wird, da das nötige Grundlagenwissen zu Master-

Data-Management-Systemen nur begrenzt ist oder zum Teil auch gänzlich fehlt [Gluchowski 2020].

Trotz dieses möglichen Lösungsansatzes und immer fortschrittlicherer ERP- und CRM-Systeme mit immer umfassenderen Funktionen steuern viele KMU ihre Stammdatenpflege noch immer manuell bzw. versuchen, das Problem in einem Tabellenkalkulationsprogramm unter Kontrolle zu bekommen.

Tatsächlich verfügen KMU weder über die Wartungsstruktur noch über das Personal, das die Werkzeuge bedienen kann, um ihre Stammdaten auf der Grundlage standardisierter Regeln und klar definierter Verantwortlichkeiten schnell und zuverlässig anzureichern.

Daher ist es auch wenig überraschend, dass die Ursachen für Datenqualitäts- und Governance-Probleme hauptsächlich in zeitaufwendigen und fehleranfälligen manuellen Datenpflegeprozessen liegen. Diese Herausforderung lässt sich nicht eingrenzen, sondern betrifft sämtliche Sektoren, Branchen, Geschäftseinheiten und Abteilungen in jeglichen Unternehmungen.

Neben der Vielzahl an Datenquellen und den darin enthaltenen Stammdaten ist auch die Datenqualität in den meisten Unternehmen eher suboptimal. Die Pflege der Informationen wird oftmals mit einfachen, aber personalintensiven und fehlerträchtigen Mechanismen durchgeführt. Intelligente Mechanismen, wie z.B. der Einsatz von EDI-Systemen oder Datenpools, sind eher selten.

Auch das Bewusstsein für eine ständige Kommunikation zwischen Fachabteilungen eines Unternehmens über die Datenqualität, die sich gegenseitig Informationen austauschen, ist oftmals nicht stark ausgeprägt. Klare Anforderungen zum Datenaustausch, Richtlinien zur Einhaltung oder auch automatische Verfahren zur Kontrolle der Datenqualität sind in vielen Unternehmen unterdimensioniert.

Die angesprochenen Herausforderungen werden mit der Errichtung eines *Data Quality Gate* angegangen. Ein *Data Quality Gate* definiert einen Vertrag zwischen Datenlieferant und Datennutzer, um ein Bewusstsein für den Austausch von Daten zu schaffen. Des Weiteren werden klare Anforderungen zum Austausch von Informationen über Fachabteilungen hinweg, Richtlinien zur Einhaltung der Datennutzung oder auch automatische Verfahren zur Kontrolle der Datenqualität definiert und gegenseitig zugesichert. Oftmals werden einheitliche Datenpools oder Systeme, die einen elektronischen Datenaustausch ermöglichen, eingesetzt [Schäffer 2017].

Abschließend ist festzuhalten, dass Stammdaten- und Datenqualitätsmanagement in einem ganzheitlichen Ansatz angegangen werden müssen, damit sowohl die Anzahl der Daten als auch deren inneren Werte, bezogen auf die Qualität und Einsatzmöglichkeiten im Unternehmen, gesteigert werden können.

14.2.2 Cloud-Infrastrukturen

Das Thema *Cloud-Infrastrukturen* ist in den letzten Jahren ein fester Bestandteil in Großunternehmen geworden. Nicht nur als Kostensenkungsfaktor in betriebs-

wirtschaftlicher Hinsicht, sondern auch als technische Weiterentwicklung in der Bereitstellung von virtuellen Arbeitsplätzen in eigens dafür aufgebauten und bereitgestellten Rechenzentren, um den Anforderungen aus den Fachbereichen gerecht zu werden.

Besonders in Großunternehmen haben sich bestimmte Formen von Cloud-Infrastrukturen etabliert, die nach der Definition des NIST [Mell & Grance 2011] wie folgt zu beschreiben sind (vgl. Abb. 14–1):

Die *Private Cloud* entspricht einer ausschließlich im eigenen Unternehmen betriebenen Cloud, die über das Intranet erreicht werden kann. Der Kunde hat dabei die volle Kontrolle. Er definiert selbst, welche Mitarbeiter oder Geschäftspartner Zugriff auf seine Daten haben.

Im Gegensatz dazu steht die *Public Cloud*. Sie wird von einem externen Dienstleister gestellt, betrieben und der Zugriff erfolgt über das Internet. Oft teilen sich in dieser Form mehrere Kunden die angebotene IT-Infrastruktur eines Anbieters und haben somit keinen direkten Einfluss auf die physische Speicherung der Daten sowie auf Rechts- und Sicherheitsaspekte. Ein wesentlicher Vorteil ist, dass der Kunde selbst entscheidet, welche und wie viele Leistungen er beziehen möchte. Dies ermöglicht ein flexibles und nutzungsbezogenes Kostenmodell (z.B. »Pay as you go«), bei dem nur das bezahlt wird, was auch wirklich benötigt wird. Dadurch sinken die Initialkosten im direkten Vergleich zu klassischen IT-Projekten. Der Nachteil liegt darin, dass eine Planung der Kosten schwieriger ist.

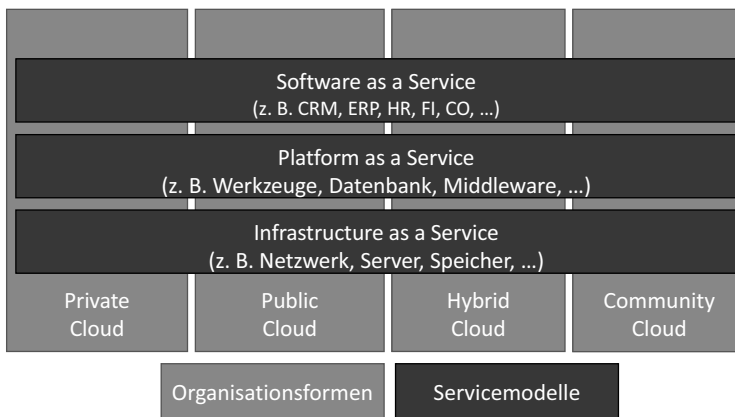


Abb. 14–1 Organisationsformen und Servicemodelle für Cloud Computing (in Anlehnung an [Pelzl et al. 2014])

Eine Spezialform des Cloud Computing, die auf der Public Cloud basiert, ist die *Community Cloud*. Sie zeichnet sich dadurch aus, dass eine fest definierte Anzahl von verschiedenen Kunden dieselben Ressourcen einer IT-Infrastruktur nutzen. Dieses Modell wird oft bei Kooperationen zwischen Unternehmen eingesetzt, um

Informationen schnell und kostengünstig auszutauschen, da die Kosten auf die verschiedenen Kooperationspartner aufgeteilt werden.

Wie so oft gibt es auch in diesem Fall nicht die eine optimale Lösung. Daher werden auch in Zukunft *Hybrid Clouds* sehr gefragt sein. Hier wird versucht, die verschiedenen Formen wie Private und Public Cloud sowie die eigene IT-Infrastruktur miteinander zu verbinden. Dabei ist der Erfolg unter anderem davon abhängig, wie gut diese verschiedenartigen Formen für den Nutzer unbemerkt als eine einzige Plattform dargestellt werden und wie leicht sich die Cloud in die eigene Infrastruktur einbinden lässt.

Im Kontrast dazu stehen die KMU vor der Herausforderung, die ersten Schritte in Bezug auf die Möglichkeiten des Cloud Computing zu gehen. Ein wesentlicher Vorteil ist hier, dass die KMU oftmals agiler und pragmatischer vorgehen als ihre großen Marktbegleiter.

Integrationsfähigkeit und Kompatibilität

Wie bereits im vorherigen Abschnitt angemerkt, stellen Hybrid Clouds den gebräuchlichsten Anwendungsfall dar. Dementsprechend ist es nachvollziehbar, dass auch kleine und mittlere Unternehmen die Integrationsfähigkeit einer Cloud-Lösung in die eigene Inhouse-Struktur zu den wichtigsten Kriterien zählen.

Besonders mittelständische Unternehmen wollen nicht ihre komplette IT-Infrastruktur in die Cloud-Lösung des Anbieters verlagern, sondern um sinnvolle Dienstleistungen ergänzen und gewinnbringend integrieren.

Somit wird auch zwangswise eine gewisse Kompatibilität zu eigenen oder externen Produkten notwendig. Je nach Cloud-Servicemodell sind verschiedene Technologien und Werkzeuge notwendig, um eine reibungsfreie Integration zu ermöglichen.

Skalierbarkeit und Vertragsindividualität

Die *Skalierbarkeit* und *Vertragsindividualität* der Services ist eines der Kernmerkmale im Cloud Computing und stellt somit eine grundlegende Voraussetzung dar, die ein Cloud-Anbieter heutzutage liefern muss. Die vereinbarten Leistungen müssen flexibel gestaltet und auf die Größe des Unternehmens angepasst sein. Ebenso müssen Support-Leistungen unter Berücksichtigung von vordefinierten Service Level Agreements bereitgestellt werden und den Vorstellungen der Kunden entsprechen.

In diesen Service Level Agreements werden die Anforderungen und Pflichten in fest vereinbarten Begriffsbestimmungen zwischen den Vertragspartnern schriftlich festgehalten, um für beide eine einheitliche begriffliche Semantik zu schaffen. Damit wird ein gemeinsamer Rahmen geschaffen, um potenzielle Unklarheiten vorab zu beseitigen. Die folgende Aufzählung zeigt beispielhaft auf, welche Punkte in einer Abstimmung zu einem SLA aufgenommen werden:

- Verfügbarkeit der bereitgestellten Leistungen
- Ausfallsicherheit und Reaktionszeiten im Ernstfall
- Szenarien zur Datensicherung und -wiederherstellung
- Eskalationswege zwischen Kunde und Dienstleister

Der Cloud-Anbieter muss dem Kunden sowohl vertraglich als auch technisch zusichern, dass er in der Lage ist, auch kurzfristige Anpassungen, z.B. Failover-Cluster-Szenarien, durchzuführen und kundenorientiert zu behandeln.

Kosteneffizienz und -transparenz

Wie bereits zu Beginn dieses Abschnitts kurz angemerkt, spielt die potenzielle Kostensenkung eine wichtige Rolle in Bezug auf die Nutzung von Cloud-Services. Dabei ist die Art des Servicemodells nachgelagert zu betrachten. Im Vordergrund steht, dass der Nutzen die Kosten übertreffen muss, im Idealfall sogar die bisher notwendigen Investitionen in die eigene Infrastruktur oder Software senkt oder gar obsolet werden lässt. Der Kostenvorteil sollte dabei die Amortisationsdauer zeitlich gesehen übertreffen. Für diese Berechnungen bieten diverse Cloud-Anbieter sogenannte Amortisationsrechner an, um Szenarien für Cloud-Infrastrukturen zu simulieren. Einsparungen lassen sich oft durch nutzungsabhängige Verrechnung, sogenannte »Pay as you go«-Modelle erwirken. Des Weiteren existieren auch Verrechnungsmodelle mit einer festen Laufzeit bzw. festem Preis, die oftmals auch noch mit einem gewissen Rabatt angeboten werden. Beide Modelle sind sowohl für den Kunden als auch für den Anbieter fair.

Neben den kalkulatorischen Bezugsgrößen ist auch die Transparenz der Kosten zu berücksichtigen. Die Cloud-Anbieter müssen die Kosten übersichtlich und nachvollziehbar darstellen. Dabei muss deutlich erkennbar sein, woraus sich die einzelnen Module zusammensetzen, was der Ursprung der Kostentreiber ist und welche Auswirkungen das auf das mögliche Kostenbudget eines Unternehmens hat.

Lage des Rechenzentrums und Datenschutz

In den letzten Jahren ist das Thema *Datenschutz* und die Lage des Rechenzentrums ein wesentlicher Faktor für die Auswahl eines Cloud-Anbieters geworden.

Durch die Lage wird der Datenfluss regional eingeschränkt. Damit sollen die Hürden für die Nutzung einer Cloud verringert werden, da die rechtliche Situation klarer erscheint. Die EU hat mit der Richtlinie 95/46/EG [Europäisches Parlament 1995] und der daraus weiterentwickelten EU-Datenschutz-Grundverordnung gewisse Mindestanforderungen an den Datenschutz und die *Datensicherheit* eingeführt, um die Verarbeitung von Daten zu gewährleisten.

Allerdings haben immer wiederkehrende Ereignisse rund um die Spionageaktionen weltweiter Geheimdienste gezeigt, dass es eigentlich keine große Rolle spielt, wo die Daten schlussendlich persistiert werden. Die Pflicht liegt beim Cloud-