Handbuch IT- und Datenschutzrecht

Auer-Reinsdorff / Conrad

3. Auflage 2019
ISBN 978-3-406-72177-9
C.H.BECK

schnell und portofrei erhältlich bei
beck-shop.de

Die Online-Fachbuchhandlung beck-shop.de steht für Kompetenz aus Tradition. Sie gründet auf über 250 Jahre juristische Fachbuch-Erfahrung durch die Verlage C.H.BECK und Franz Vahlen.
beck-shop.de hält Fachinformationen in allen gängigen Medienformaten bereit: über 12 Millionen Bücher, eBooks, Loseblattwerke, Zeitschriften, DVDs, Online-Datenbanken und Seminare. Besonders geschätzt wird beck-shop.de für sein umfassendes Spezialsortiment im Bereich Recht, Steuern und Wirtschaft mit rund 700.000 lieferbaren Fachbuchtiteln.

kleinen Datenmengen getestet, um in der Entwicklung schnell vorankommen zu können. Bei ersten Tests mit der kompletten Menge an zu migrierenden Daten stellt sich jedoch dann schnell heraus, dass frühere Zeitschätzungen den tatsächlichen Zeitbedarf grob unterschätzt haben.

(4) Klärung aller Rahmenbedingungen. Die Migration von Daten von einem oder mehreren Anwendungssystemen in ein anderes Anwendungssystem ist nicht nur eine technische Aufgabe, sondern impliziert auch stets die Klärung betrieblicher, fachlicher, wirtschaftlicher und juristischer Fragen. So ist zum Beispiel zu klären, ob offene Bestellungen noch in den Altsystemen abgewickelt werden sollen, wenn sich das Bestellwesen insgesamt aufgrund des Einsatzes eines Neusystems ändert. Auch neue oder geänderte Gesetze können die Anforderungen an eine Datenmigration beeinflussen. **149**

bb) Technische Fragestellungen. (1) Struktur, Qualität und Inhalte der zu migrierenden Daten. Viele Auftraggeber gehen davon aus, dass ihre Altdaten gut strukturiert sind, ausreichend dokumentiert vorliegen und in Bezug auf die Qualität als mindestens „brauchbar" einzustufen sind. In der Praxis ist jedoch die Erfahrung zu machen, dass die Struktur und die Qualität zu migrierender Altdaten in nahezu allen Fällen als mangelhaft einzustufen ist. **150**

Zu den häufigsten Mängeln der Altdaten gehören: **151**

- Fehlende Dokumentation der Struktur der Daten
- Duplikate in den Datensätzen
- Inkonsistente Daten
- Unvollständige Datensätze/fehlende Belegungen für Datenfelder
- Zweckentfremdete Nutzung von Datenfeldern[50]
- Falsche/unvollständige/fehlende Referenzen zwischen zusammengehörigen Daten
- Inhaltlich falsche Daten

Solche Mängel können in IT-Systemen aufgrund von fehlenden oder fehlerhaften Funktionalitäten, wegen Nichtbenutzung bestimmter Funktionalitäten oder aufgrund von fehlenden Benutzerberechtigungen jahrelang unerkannt bleiben. Es kommt hinzu, dass die maßgeblichen Datenmodelle zumeist nicht dokumentiert sind und dass die internen Fachkräfte, die die Struktur und die Inhalte der Daten kennen, in aller Regel äußerst rar oder manchmal auch gar nicht mehr vorhanden sind. Der Auftragnehmer/IT-Dienstleister kann sich also nicht darauf verlassen, dass er vorhandene Altdaten seines Auftraggebers ohne intensive Prüfung in das von ihm entwickelte Neusystem übernehmen kann.[51] In jedem Fall bleibt jedoch der Auftraggeber für die Qualität und Inhalte seiner Altdaten verantwortlich.[52] **152**

Um die Analyse aller Altdaten zu vereinfachen, werden die Altdaten aus verschiedenen Quellen meistens in ein einziges Repository übertragen, wo sie dann unter Berücksichtigung von allen existierenden Querbeziehungen analysiert werden können. **153**

(2) Umfang der zu migrierenden Daten. In der Praxis ist häufig wahrzunehmen, dass die Vertragspartner über den Umfang der zu migrierenden Daten nur sehr ungenaue oder manchmal sogar gar keine Vereinbarungen getroffen haben. In solchen Fällen wären dann während der Laufzeit des Projekts geeignete Vereinbarungen zu treffen, was aber nicht selten zu Streit zwischen den Vertragspartnern führt. Gelingt es den Vertragspartnern nicht, die notwendigen Details im Hinblick auf den Umfang der zu migrierenden Daten festzulegen, bleibt (weitgehend) unklar, in welcher Breite und Tiefe Altdaten zu migrieren sind[53] und welchen Beitrag die Vertragspartner dabei im Einzelnen zu leisten haben. **154**

[50] Zum Beispiel werden in dem Feld „Telefonnummer" auch gerne Namen von Ansprechpartnern gespeichert. Bei einer Datenmigration wäre dann zu berücksichtigen, dass die Telefonnummern von den Namen der Ansprechpartner technisch getrennt werden, jedoch über interne technische Referenzen als zusammengehörig gekennzeichnet werden.

[51] Aus rechtlicher Sicht ist aus diesem Grunde zu diskutieren, inwieweit eine Prüfungspflicht der Daten vor der Datenübernahme besteht.

[52] Siehe hierzu *Schneider* M. Rn. 336.

[53] Ein Beispiel für die „Breite" wäre die Entscheidung, dass im Kontext eines neuen ERP-Systems Lieferanten migriert werden, während sich die „Tiefe" darüber bestimmen würde, dass nur solche Lieferanten berücksichtigt werden, von denen zum Beispiel in den letzten drei Jahren Waren gekauft worden sind.

beck-shop.de DIE FACHBUCHHANDLUNG

155 Aus Sicht des Kunden bzw. Auftraggebers sind in nahezu allen Fällen immer alle erdenklichen Daten zu migrieren, aus Sicht des IT-Dienstleisters erscheint dies jedoch oft nicht erforderlich, selbst wenn dies bedeutet, dass ein oder mehrere Altsysteme zum Zwecke des Zugriffs auf historische Daten weiterbetrieben werden müssen. Um die Komplexität der Datenmigration zu senken, macht es zum Beispiel Sinn, offene Bestellungen noch im Altsystem abzuwickeln, um dann nur die „Ergebnisse" der Transaktionen in das neue System übernehmen zu müssen. Was aber letztlich wirtschaftlich und fachlich sinnvoll ist, muss der Auftraggeber/Kunde verantworten.

156 In aller Regel ist die Frage des Umfangs der zu migrierenden Daten in Bezug auf Breite und Tiefe eine aus fachlicher Sicht höchst komplexe Fragestellung. Enthält der Projektvertrag hierzu keine Vereinbarungen, ist der Streit zwischen den Vertragspartnern meistens vorprogrammiert – auf der Basis des mittleren Ausführungsstandards und dessen, was der Kunde erwarten darf, eventuell in Kombination mit dem, was üblich ist und was dem Gebrauch des gelieferten Werks entspricht, ist der Umfang einer Datenmigration auch von erfahrenen EDV-Sachverständigen nur äußerst schwer zu bestimmen. Die Vertragspartner gehen also ein hohes Risiko ein, wenn der Umfang der Datenmigration nicht im Detail festgelegt ist.

157 *(3) Datentransformation.* Die Vertragspartner müssen davon auszugehen, dass die Daten, die den Altsystemen entnommen werden, vor der Übertragung in das Neusystem entsprechend aufbereitet bzw. transformiert werden müssen. Diese Aufbereitungs- bzw. Transformationsschritte können fachlich und technisch äußerst komplex sein, da das Neusystem die Daten in aller Regel nicht nur in neuen Formaten und in einer neuen Struktur verwaltet, sondern auch zum Teil neue bzw. andere Funktionalitäten umfasst.

158 Typische Probleme bei der Transformation von Altdaten sind:

- Formatkonvertierungen (Datumsangaben, Telefonnummern, Bestellnummern etc.)
- Auftrennung von Datenfeldern (bei einem Kombinationsfeld Name/Titel, bei Produktbezeichnungen etc.)
- Zusammenfassung von Datenfeldern (zB für Texte in einem Produktkatalog)
- Ergänzung von Pflichtfeldern mit Standardbelegungen, die in den Altsystemen nicht vorgesehen sind (zB Bonität eines Kunden)
- Fehlende Zuordnung von Datensätzen verschiedener Datenbanktabellen

beck-shop.de DIE FACHBUCHHANDLUNG

159 Die Vertragspartner sind vor diesem Hintergrund gut beraten, die Verantwortlichkeiten für die erforderliche Transformation der Daten vertraglich im Detail festzulegen oder zumindest während der Laufzeit des Projekts verbindlich zu vereinbaren. In aller Regel besitzt der IT-Dienstleister entsprechende Werkzeuge, um vorhandene Altdaten in die vom Neusystem vorausgesetzte Struktur zu überführen; gelegentlich müssen aber auch individuelle Programme entwickelt werden, um die erforderlichen Datentransformationen durchführen zu können.

160 *(4) Einspielung von Altdaten in ein Neusystem.* Nachdem die fachlich notwendigen Altdaten aus den betroffenen Altsystemen extrahiert worden sind und die entsprechenden Datentransformationen durchgeführt worden sind, müssen die aufbereiteten Altdaten in das betroffene Neusystem eingespielt werden. Hierzu gibt es drei wesentliche technische Mechanismen, die alternativ oder auch in Kombination zur Anwendung kommen können:

1. Import der Altdaten via Dateischnittstelle
2. Übertragung der Altdaten via Programmierschnittstelle
3. Übertragung mit speziellen Werkzeugen, die einen Datentransfer inkl. Transformation in einem Schritt durchführen

161 Während Alternative 1) zumeist rasch zu realisieren ist, bietet die Alternative 2) bessere Möglichkeiten, um zum einen Fehlerfälle zu behandeln und um zum anderen die Konsistenz der Daten im Neusystem sicherzustellen. Alternative 3) bietet den wesentlichen Vorteil, dass auf eine vorhandene Lösung zurückgegriffen werden kann, die auch die notwendigen Datentransformationen mit erledigt.

162 *(5) Schnittstellen zu Nachbarsystemen.* Wenn im Zuge der Veränderung einer Systemlandschaft Altsysteme mit Schnittstellen zu Nachbarsystemen abgelöst werden sollen, ist zu

berücksichtigen, dass auch in den meisten Fällen Stammdaten der Nachbarsysteme in das betroffene Neusystem zu migrieren sind.

Daneben ist zu berücksichtigen, dass eine vorhandene Schnittstelle zu einem Nachbarsystem entweder unverändert bleibt, was bedeutet, dass die bisherige Spezifikation der Schnittstelle weiterhin ihre Gültigkeit behält, oder die Schnittstelle wird geändert, was dann nach sich zieht, dass sowohl das Nachbarsystem als auch das Neusystem an die neue bzw. geänderte Schnittstellenspezifikation angepasst werden müssen. Im Hinblick auf die Migration von Daten wird zu prüfen sein, ob Daten aus den Nachbarsystemen – und wenn ja, welche – in das neue System zu übernehmen sind (oder sogar umgekehrt). **163**

(6) Fehlende und „überflüssige" Daten. Da ein neues IT-System in aller Regel gegenüber vorhandenen Altsystemen neue Funktionalitäten mit sich bringt, ist es sehr wahrscheinlich, dass die Übertragung von Daten aus den Altsystemen alleine nicht ausreicht, um einen sinnvollen Betrieb des neuen Systems zu gewährleisten; die im Neusystem „fehlenden" Daten müssen entweder manuell ergänzt oder aus anderen Quellen in das neue System eingespielt werden. **164**

Umgekehrt kann es auch vorkommen, dass die Altsysteme Daten beinhalten, die im Neusystem nicht verarbeitet werden können, zumindest nicht in der Version des Neusystems, die produktiv gesetzt werden soll. In diesem Fall ist der Auftraggeber gefordert, sich Gedanken darüber zu machen, ob die betroffenen überflüssigen Daten überhaupt nicht mehr im Rahmen seines Geschäftsbetriebs genutzt werden sollen oder ob die Daten mit Hilfe der Altsysteme oder eines eigens angelegten Archivs einsehbar bleiben müssen. **165**

cc) Durchführung der Migration und Qualitätssicherungsmaßnahmen. (1) Personelle Besetzung des Datenmigrations-Teams. In der Praxis ist darauf zu achten, dass das Projektteam für eine Datenmigration mit den „richtigen" Fachkräften besetzt wird – dies stellt hohe Anforderungen sowohl an den Auftraggeber, als auch an den Auftragnehmer (IT-Dienstleister). Zu den wichtigsten Anforderungen in Bezug auf die personelle Besetzung eines geeigneten Teams gehören folgende Punkte: **166**

beck-shop.de DIE FACHBUCHHANDLUNG

Nr.	Anforderung	Auftraggeber	Auftragnehmer
1	Inhaltliche und strukturelle Analyse der Altdaten	X	
2	Festlegung der fachlichen Anforderungen für die Datenmigration inklusive Festlegung des Austauschformats für die Daten	X	
3	Technische Konzeption der Datenmigration		X
4	Tiefgehende Kenntnisse über Datenmigrationswerkzeuge		X
5	Programmierkenntnisse bzgl. der Entwicklung neuer Werkzeuge		X
6	Abnahme der Datenmigration[54]	X	

In der Praxis ist häufig zu beobachten, das der Auftraggeber für seine Altdaten praktisch keine verwertbare Dokumentation vorweisen kann und auch nicht (mehr) über das Personal verfügt, das über die Struktur und die Inhalte der Daten präzise Auskunft geben könnte. Die meisten Auftraggeber haben daher wenig Überblick über die Daten in ihren Altsystemen, woraus letztlich zu folgern ist, dass sich der Auftraggeber bei der Analyse der Altsysteme und der fachlichen Konzeption der Datenmigration gegebenenfalls durch externe Kräfte unterstützen lassen muss.[55] **167**

[54] Hierbei wird angenommen, dass die Lieferungen und Leistungen eines Auftragnehmers in Bezug auf eine Datenmigration in den meisten Fällen wohl als werkvertragliche Leistungen eingeordnet werden – so zumindest auch *Schneider* M. Rn. 337.

[55] Für die hier erwähnten externen Kräfte kommt in aller Regel nicht die Implementierungsfirma des Neusystems in Frage, sondern eher Spezialisten, die mit der Fachlichkeit und der Technologie der abzulösenden Altsysteme vertraut sind.

168 Ein typisches Datenmigrationsteam würde sich in der Praxis etwa folgendermaßen zusammensetzen:[56]

Nr.	Rolle	Auftraggeber	Auftragnehmer
1	Projektmanager		X
2	Koordinator	X	
3	Fachbereichsmitarbeiter	X	
4	IT-Mitarbeiter	X	
5	Berater		X
6	Entwickler		X

169 *(2) Qualitätsmanagement.* Eine Altdatenübernahme bzw. eine Datenmigration ist in einem komplexen Umfeld eines großen Unternehmens oder einer Behörde stets eine große Herausforderung. Die einzelnen Verfahrensschritte sind schon für sich betrachtet komplex und fehleranfällig. Daraus ergibt sich unmittelbar, dass ein sachgerechtes Qualitätsmanagement etabliert werden muss, was nicht nur bedeutet, dass die Qualität bzw. das Ergebnis der Datenmigration am Ende der Verfahrensschritte zu überprüfen ist, sondern dass während des gesamten Spezifikations-, Entwicklungs-, Integrations- und Durchführungsprozesses bereits in kleinen Schritten zu überprüfen und zu kontrollieren ist, ob die jeweiligen Qualitätsziele erreicht werden („prozessbegleitendes Qualitätsmanagement"). Dies bedeutet auch, dass eine entsprechende Dokumentation gepflegt sein muss und dass die Zwischenergebnisse der Datenmigration tatsächlich auch auswertbar sind.[57]

170 Jeder Auftraggeber sollte daher darauf drängen, dass sein Auftragnehmer/IT-Dienstleister seinerseits entsprechende Maßnahmen im Sinne des hier dargestellten Qualitätsmanagements ergreift. Umgekehrt wird auch jeder Auftragnehmer ein hohes Interesse daran haben, dass die vom Auftraggeber zu erbringenden Leistungen von einem angemessenen Qualitätsmanagement erfasst werden.

beck-shop.de DIE FACHBUCHHANDLUNG

171 *(3) (Abnahme-)Test der Ergebnisse einer Datenmigration.* Nach einer erfolgten Datenmigration stellt sich für den Auftraggeber oft die Frage, wie die vom Auftragnehmer erbrachten Lieferungen und Leistungen getestet bzw. abgenommen werden können.

172 In aller Regel wird es sich empfehlen, zunächst die Vollständigkeit und Korrektheit der bereitgestellten Altdaten zu überprüfen, da sich ansonsten schwer zu findende Folgefehler ergeben können.[58]

173 Wenn die Altdaten im Neusystem angekommen sind, besteht für den Auftraggeber die Möglichkeit, den Erfolg der Datenmigration dadurch zu prüfen, dass passende Testfälle im Neusystem durchgespielt und überprüft werden, ob alle Daten korrekt verarbeitet werden. Sollte sich herausstellen, dass eine tiefergehende Analyse der Daten erforderlich ist, so müssen zusätzlich die Inhalte von Datenbanken des Neusystems geprüft werden, was jedoch wiederum die Unterstützung des Auftragnehmers erfordern kann.

174 Auch zu diesem Thema muss den Vertragspartnern empfohlen werden, die Einzelheiten des (Abnahme-)Tests vertraglich zu regeln, damit insbesondere die Verantwortlichkeiten zwischen den beteiligten Vertragspartnern klar definiert sind.

175 **d) Zusammenfassung.** Um die Übernahme von Altdaten in ein neues IT-System gewährleisten zu können, müssen die Vertragspartner in aller Regel eine beträchtliche Menge an betrieblichen, organisatorischen, fachlichen, technischen und wirtschaftlichen Fragen klären. Die Klärung dieser Fragen ist einerseits vor dem Hintergrund wichtig, dass für die Migration der betroffenen Daten eine Planung bzw. ein Aktivitäten- und Fristenplan erstellt werden kann, andererseits müssen die Vertragspartner auch genau wissen, welche Aufgaben in

[56] Siehe auch das White-Paper von BSM, Abschnitt 3.2 („Data Migration Teams").
[57] Vergleiche hierzu auch das White-Paper von BSM, Abschnitt 3.5 („Planning for Audit").
[58] Die Überprüfung der Altdaten sollte ohnehin Bestandteil der Qualitätssicherungsmaßnahmen des Auftraggebers sein.

ihren Verantwortungsbereich fallen, damit diese Aufgaben qualitäts- und termingerecht erledigt werden können.

Da heutige Systemlandschaften in großen Unternehmen und Behörden zum Teil extrem komplex sind, ist auch zu beachten, dass in den meisten Fällen Nachbarsysteme beteiligt sein werden, die die Übernahme von Daten verkomplizieren. In jedem Fall sollten die Vertragspartner alle erforderlichen Details der Datenübernahme klären und die Datenübernahmeprogramme entsprechend darauf ausrichten. Obwohl es in vielen Fällen möglich ist, vorhandene Standardwerkzeuge für die Migration von Daten einzusetzen, darf nicht übersehen werden, dass abhängig von der Aufgabenstellung im Einzelfall auch individuelle Datenmigrationsprogramme zu entwickeln sind, was typischerweise sehr aufwendig und zeitintensiv ist. 176

Den Vertragspartnern muss auf jeden Fall davon abgeraten werden, die Klärung von notwendigen Details der Datenmigration auf eine zu späte Projektphase zu verschieben, da die Komplexität einer Datenmigration durchaus die Größenordnung des eigentlichen Entwicklungsprojekts für das Neusystem erreichen kann, dh in schwierigen Fällen kann die Projektdauer durchaus 6–24 Monate betragen. 177

5. Datenkonsistenz

a) Einführung und Begrifflichkeiten. Unter dem Begriff „Datenkonsistenz" (in Informatikkreisen auch gelegentlich als „Datenintegrität" bezeichnet) versteht man, dass eine Menge von vorliegenden Daten fachlich und technisch zusammenpassen und in Bezug auf die Querbeziehungen untereinander keine Unstimmigkeiten aufweisen. 178

Die Konsistenz von Daten spielt in aller Regel eine Rolle im Kontext von Datenbanken, die über entsprechende Mechanismen sicherstellen können, dass konsistente Daten auch bei Änderungen immer wieder in konsistente Daten überführt werden können (die Datenkonsistenz bleibt dadurch immer gewährleistet bzw. erhalten).[59] Bei einer falschen oder unsachgerechten Nutzung einer Datenbank kann die Konsistenz der Daten jedoch verletzt werden, was Fehler oder Abstürze der mit der Datenbank verbundenen Anwendung nach sich ziehen kann. Umgekehrt lässt sich sagen, dass eine korrekt programmierte Software mit einer darunterliegenden Datenbank die Konsistenz der zu verwaltenden Daten stets bewahrt, es sei denn, dass die Datenbanksoftware selbst (also die Software des Datenbankherstellers) Fehler aufweist. 179

beck-shop.de
DIE FACHBUCHHANDLUNG

Bei der Spezifikation von Softwaresystemen wird die Anforderung, dass die zu verwaltenden Daten stets konsistent sein müssen, nur selten explizit aufgelistet, da jeder Auftraggeber davon ausgeht und auch davon ausgehen kann, dass das Softwareentwicklungshaus das beauftragte Softwaresystem so entwickelt, dass die Datenkonsistenz stets gewährleistet ist. Die Anforderung bezüglich der Datenkonsistenz kann somit als implizite Qualitätsanforderungen angesehen werden. 180

b) Referenzierte Integrität. In einer Datenbanktabelle T1 identifizieren die Attributwerte eines Schlüssels einen Datensatz D1 eindeutig. Hierdurch ist es möglich, von einer anderen Datenbanktabelle T2 aus auf einen bestimmten Datensatz der ersten Datenbanktabelle (T1) zu verweisen, zum Beispiel auf den Datensatz D1. Im Kontext der Datenbanktabelle T2 wird der Schlüssel, der als Verweis auf den Datensatz D1 in der Tabelle T1 dient, auch als „Fremdschlüssel" bezeichnet. 181

Wenn genannten Beispiel der Datensatz D1 in der Tabelle T1 gelöscht wird, wäre die referenzierte Integrität verletzt, da der Fremdschlüssel des Datensatzes D2 in der Tabelle T2 nicht mehr auf einen gültigen Datensatz zeigt. Ein solcher, nicht mehr gültiger Fremdschlüssel wird in der Informatik auch als „dangling reference" bezeichnet. 182

In einer Datenbank kann die referenzierte Integrität durch entsprechende Programmierung stets sichergestellt werden.

c) Komplexe Integritätsbedingungen. In der Praxis kommt es häufig vor, dass die referentielle Integrität über verschiedene Datenbanktabellen hinweg sichergestellt werden muss. 183

[59] Siehe hierzu ausführlich *Kemper/Eickler*: Datenbanksysteme – eine Einführung, 9. Auflage, Kap. 5 (Datenintegrität und temporale Daten), S. 163 ff.

So könnte zum Beispiel in einer Anwendung gefordert sein, dass eine Bestellung mindestens eine Artikelposition enthalten muss und dass jedem Artikel mindestens eine oder mehrere mögliche Verpackungen zugeordnet sind.

Durch eine entsprechende Anwendungsprogrammierung lassen sich auch komplexe Integritätsbedingungen sicherstellen.

184 **d) Trigger.** Mithilfe von Triggern wird in Datenbanken sichergestellt, dass konsistente Datenzustände hergestellt werden, wenn Änderungen an Datensätzen in Datenbanktabellen vorgenommen werden. Die Programmierung eines entsprechenden Triggers geschieht innerhalb einer SQL-Datenbank mithilfe einer speziellen Programmiersprache.

185 Mit einem Trigger könnte beispielsweise erreicht werden, dass ein Mitarbeitergehalt nur erhöht, aber nie erniedrigt werden kann. Ein entsprechender Trigger würde bei dem Vorliegen gewisser Bedingungen zuvor definierte Aktionen „loslaufen" lassen.

6. Big Data

186 **a) Einführung und Begriffsklärung.** „Big Data" ist ein Schlagwort, das einerseits große Mengen von Daten bezeichnet (sog. „Massendaten"), andererseits aber auch moderne, digitale Verfahren und Technologien, mit deren Hilfe Massendaten nutzbringend ausgewertet werden können.[60]

187 Die zu analysierenden und auszuwertenden Ausgangsdaten stammen in aller Regel aus unterschiedlichen Quellen und sind durch unterschiedliche Struktur, Schnelllebigkeit und variierende Qualität gekennzeichnet.

188 Aus Informatiksicht sind Massendaten und entsprechende Verarbeitungsmethoden nicht völlig neu, jedoch erfordern die in der Praxis vorliegenden Gegebenheiten hinsichtlich der Datenvolumina, der ständigen Änderungen an den Inhalten und der Struktur der Daten sowie der Grad der Verteilung der Daten zumeist neue Ansätze in Bezug auf die Aufbereitung, Analyse und Auswertung der Daten.

beck-shop.de DIE FACHBUCHHANDLUNG

189 Mit der starken Entwicklung der Informationstechnologie und dem enormen Preisverfall von Hardware- und Speicherkosten ist der Einsatz von Big Data Technologien für viele Unternehmen, Kunden und IT-Dienstleister hoch interessant geworden, um Wettbewerbsvorteile zu erlangen. Cloud Architekturen und Open Source Software haben sehr stark dazu beigetragen, dass ein nennenswerter Geschäftsnutzen (business value) aus den Big Data Verfahren gezogen werden kann. Die bekannten Unternehmen Google, Facebook und Amazon sind Beispiele dafür, dass mit Ideen auf der Grundlage von Big Data ein beachtliches Unternehmenswachstum und eine damit verbundene hohe Unternehmensbewertung erreicht werden können.

190 **b) Datenquellen von Big Data.** Das Ausgangsdatenmaterial für Big Data Verfahren kann äußerst vielfältig sein, zum Beispiel sind in der Praxis Chats aus sozialen Netzwerken, Web Server Protokolle, Daten von Sensoren, Bilder von Satelliten, Banktransaktionen, GPS-Daten, Live-Daten aus Fahrzeugen, Inhalte von Webseiten oder alle Arten von Dokumenten und Dateien anzutreffen. Als unmittelbare Konsequenz ergibt sich, dass die verschiedenen Datenarten sehr stark dazu beitragen, dass die Daten in ihrer Gesamtheit uneinheitlich strukturiert sind und dass entsprechende Verarbeitungsverfahren auf diesen Umstand angepasst werden müssen.

191 **c) Ergebnisse einer Big Data Verarbeitung.** Als Ergebnis einer Big Data Verarbeitung wird sich in aller Regel ergeben, dass ein Bericht (Report) mit strukturierten Daten erzeugt wird, damit auf dieser Basis Entscheidungen getroffen werden können oder um auf dem Wege einer Feedback-Loop gewonnene Daten wieder in die Ausgangsdatenbasis zurückspielen zu können.

192 Die sich ergebende Datenmenge einer Big Data Verarbeitung ist typischerweise gegenüber dem Ausgangsdatenmaterial sehr klein.

[60] Siehe hierzu zum Beispiel *Dumbill,* What is big data? An introduction to the big data landscape, 11.1.2012, https://www.oreilly.com/ideas/what-is-big-data, Abruf vom 17.5.2018.

Äußerst schwierig ist in aller Regel die Frage nach der Datenqualität der Ergebnisdaten zu beurteilen, da schon die Datenqualität des Ausgangsdatenmaterials aufgrund des großen Volumens nicht festzustellen ist.

d) Technische Aspekte. Aus technischer Sicht stellen sich im Hinblick auf Big Data neue Herausforderungen, um verteilt gespeicherte, schnelllebige, große Datenmengen in akzeptabler Zeit verarbeiten zu können. Die bedeutendsten Herausforderungen liegen in folgenden Bereichen: 193

- Da Big Data Daten typischerweise unterschiedlich strukturiert vorliegen, müssen die digitalen Verfahren zum Lesen dieser Daten mit unterschiedlichen Datenformaten und -strukturen zurechtkommen. Im Hinblick auf den Aufbau einer entsprechenden Tool-Landschaft zur Verarbeitung des Ausgangsdatenmaterials ergibt sich daher ein erheblicher Konfigurations- und Programmieraufwand.
- Aufgrund der starken Verteilung der Big Data Daten kann die Bandbreite der Übertragungswege ein Problem darstellen. Es kann daher erforderlich sein, die zu analysierenden Daten an Ort und Stelle zu filtern und zu verdichten, um die Datenvolumina, die letztlich übertragen werden sollen, klein zu halten.
- Um eine massiv parallele Verarbeitung realisieren zu können, wird es in aller Regel nötig sein, Ressourcen einer Cloud-Umgebung heranzuziehen. Auch hier wird sich das Problem geringer Bandbreiten der Übertragungswege ergeben. Nach der Verarbeitung ist dann auch die Frage zu lösen, wie nach der massiv parallelen Verarbeitung von Daten die Teilergebnisse zusammengeführt werden können.
- Die Schnelllebigkeit der Big Data Ausgangsdaten erfordern spezielle technische Lösungen, die sich den Veränderungen in Bezug auf die Inhalte und die Struktur der Daten sehr schnell anpassen können. Herkömmliche Ansätze auf der Basis von SQL und starrer Programmierung dürften den Anforderungen in aller Regel nicht mehr gerecht werden.

Es wird sich empfehlen, bei einem konkret vorliegenden Big Data Projekt möglichst genau zu analysieren, mit welchem Inhalt und mit welcher Struktur die Ausgangsdaten vorliegen, um dann in einem zweiten Schritt spezifizieren zu können, welche Ergebnisse aufgrund der Auswertungen benötigt werden. Wie im Software Engineering üblich, sollte jeder Implementierung eine fachliche und technische Spezifikation vorausgehen. 194

beck-shop.de DIE FACHBUCHHANDLUNG

II. Datenbanken

Ein **Datenbankmodell** ist das Konzept, die theoretische Grundlage, für ein Datenbanksystem und bestimmt, auf welche Art und Weise Daten prinzipiell in einem Datenbanksystem gespeichert werden und wie man die Daten manipulieren (zugreifen und ändern) kann. Nach Edgar F. Codd definiert sich ein Datenbankmodell aus drei Eigenschaften: 195

1. einer Sammlung von Datenstrukturen,
2. einer Menge von Operatoren, die auf jede der Datenstrukturen unter 1. angewandt werden kann, um Daten abzufragen oder abzuleiten,
3. einer Menge von Integritätsregeln, die implizit oder explizit Veränderungen der Daten erlauben.

Die heute wichtigsten Datenbankmodelle sind: 196

- Relationales Datenbankmodell.
- Objektorientiertes Datenbankmodell.

1. Relationales Datenbankmodell

Eine **relationale Datenbank** ist eine Datenbank, die auf dem relationalen Datenbankmodell basiert, das von Edgar F. Codd 1970 erstmals vorgeschlagen wurde; darin ist Relation ein im streng mathematischen Sinn wohldefinierter Begriff, der im Wesentlichen ein mathematisches Modell für eine Tabelle (Relation) beschreibt. Die Daten werden dabei in Form von zweidimensionalen Tabellen verwaltet, die über Schlüssel (Primärschlüssel, Fremdschlüssel) miteinander verknüpft werden können. Die meisten in der Praxis eingesetzten Datenbanksysteme sind für relationale Datenbanken konzipiert (relationale Datenbankmanagementsysteme, 197

kurz RDBMS). Im allgemeinen Sprachgebrauch ist deshalb oft eine relationale Datenbank oder ein relationales Datenbanksystem gemeint, wenn von Datenbanken die Rede ist.

198 Für relationale Datenbanken gibt es mit **SQL** eine verbreitete und größtenteils standardisierte Abfragesprache.[61] Zur Modellierung von relationalen Datenbanken wird meist das Entity-Relationship-Modell oder Varianten davon verwendet. Es dient zum Entwurf eines konzeptuellen Schemas, welches unter Verwendung eines DBMS implementiert werden kann. Dieser Schritt wird als logischer Entwurf oder auch Datenmodellabbildung bezeichnet und hat als Ergebnis ein Datenbankschema im Implementierungsdatenmodell des DBMS.

199 Früher wurden in der betrieblichen Datenverarbeitung hierarchische Datenbanksysteme und Netzwerk-Datenbanksysteme verwendet. Sie kommen in Spezialfällen auch heute noch zum Einsatz. Zum Teil werden die relationalen Datenbanken durch objektorientierte Datenbanken abgelöst. Relationale Datenbanken sind aber derzeit immer noch die am meisten verbreitete Datenbankform und es ist nicht klar, ob sich die objektorientierten Datenbanken durchsetzen werden. Die großen Datenbankhersteller fügen ihren relationalen Datenbanken objektorientierte Eigenschaften hinzu und haben so die objektrelationalen Datenbanken entwickelt.

200 Die Grundlagen der Theorie der relationalen Datenbank wurden von Edgar F. Codd in den 1960ern und 1970ern gelegt[62] und beschrieben. Theoretisch basieren alle Operationen auf der relationalen Algebra.

201 Die erste kommerziell erfolgreiche relationale Datenbank wurde jedoch erst Ende der 1970er von der Firma Oracle auf den Markt gebracht.

Die **Grundregeln für eine relationale Datenbank** (nach Codd) lassen sich wie folgt beschreiben:

- Jede Relation ist eine zweidimensionale Tabelle und entspricht einem Relationstyp.
- Jede Zeile dieser Relation (Tabelle) wird Tupel genannt und beschreibt ein konkretes Tupel des Relationstyps, den die Relation (Tabelle) darstellt.
- Jede Spalte der Relation (Tabelle) entspricht einem Attribut des Relationstyps. Die konkreten Tupel werden somit durch die entsprechenden Attributwerte beschrieben.
- Der Grad einer Relation bezeichnet die Anzahl der Attribute.
- Die Kardinalität einer Relation entspricht der Anzahl der Tupel.
- Existiert für ein Attribut eine begrenzte Anzahl von Attributwerten, so wird die Zusammenfassung aller Attributwerte für dieses Attribut Domäne (Wertebereich) genannt.
- Es ist nicht relevant, in welcher Reihenfolge Zeilen bzw. Spalten der Tabelle angeordnet sind.
- Die Existenz zweier identischer Zeilen ist ungültig.
- Attributwerte sind atomar.

beck-shop.de DIE FACHBUCHHANDLUNG

Wichtige Begriffe relationaler Datenbankmodelle

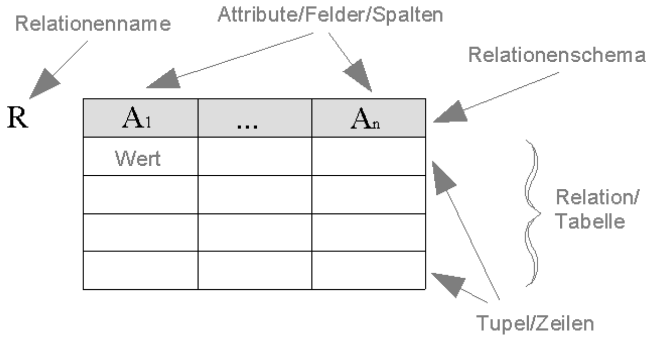

Abbildung 4.3a: Begriffe relationaler Datenbankmodelle

[61] Einige der wichtigsten bzw. bekanntesten relationalen Datenbanken sind: *Oracle*, MS SQL Server, MS Access, Informix, INGRES, TransBase, mySQL, postgres, IBM DB2.

[62] *Codd*, A relational model of data for large shared data banks, 1970, ISSN 0001–0782 oder auch: Codd, The relational model for database management, 1990, ISSN 0–201-14192-2.