

Anhang C: Detaillierte Beschreibung der Schlüsselkonzepte des Datenmanagements

C.1 Datenunabhängigkeit

Entwurfsprinzip

„Der Definition zufolge beschreibt Datenunabhängigkeit die Möglichkeit, das Schema auf einer Ebene ändern, ohne das Schema der nächsthöheren Ebene ändern zu müssen.“ (Elmasri und Navathe, 2009)

Es wird zwischen logischer und physischer Datenunabhängigkeit unterschieden: Die Erste befasst sich mit Änderungen des konzeptionellen Schemas und thematisiert somit das Hinzufügen, Verändern oder Entfernen von Datenfeldern oder Datensatztypen. Die Zweite berücksichtigt hingegen die Möglichkeit, die physischen Daten neu zu organisieren, ohne das konzeptionelle Schema ändern zu müssen.

Datenunabhängigkeit ist daher eine Folge der klaren Einteilung eines Systems in mehrere konzeptionelle Schichten.

Kernaussagen

- Die Benutzungsoberfläche von Datenmanagementsystemen abstrahiert von der internen Speicherung.
- Die Sichtbarkeit von internen Aspekten wird je nach Benutzerrolle eingeschränkt.
- Zur Trennung des Systems in Schichten werden klare Schnittstellen definiert.
- Datenmanagementsysteme erlauben es, einzelne Schichten auszutauschen, ohne andere anzupassen.
- Insbesondere kann die konzeptionelle Schicht (das Datenmodell) flexibel an sich ändernde Anforderungen angepasst werden.
- Informatiksysteme abstrahieren interne Details, indem diese geeignet versteckt werden, mit dem Ziel Systeme besser nutzbar zu machen.
- Die entstehenden Schichtenarchitekturen werden in der Informatik verbreitet eingesetzt.
- Jede Benutzergruppe hat unterschiedliche Anforderungen, denen ein System genügen muss.

- Für eine Benutzergruppe irrelevante Informationen und Möglichkeiten sollten erkannt und verborgen werden (Geheimnisprinzip).
- Geeignete Schnittstellen müssen (nicht nur bei der Teilung eines Systems in mehrere Schichten) klar definiert werden.

Verwandte Konzepte des Datenmanagements

- Strukturierung
- Repräsentation
- Dauerhaftigkeit
- Verfügbarkeit
- Partitionierung

Relevanz in Verbindung zu Praktiken des Datenmanagements

Datenunabhängigkeit ist insbesondere im Bereich der Modellierung, Implementierung, Optimierung und Analyse relevant, da sich diese Praktiken konkret mit der Strukturierung und Speicherung von sowie dem Zugriff auf Daten beschäftigen.

Anknüpfungspunkte in der Informatik

- **Schichtenarchitekturen:** Die Trennung eines Systems in verschiedene architekturelle Schichten mit verschiedenen Aufgaben kommt in verschiedensten Bereichen der Informatik vor, beispielsweise in der Softwareentwicklung, der Netzwerkkommunikation oder bei Betriebssystemen.
- **Schnittstellendefinitionen:** Die Definition von Schnittstellen ist insbesondere auch bei der Modularisierung von Anwendungen, aber auch für die Netzwerkkommunikation oder den Zugriff auf externe Datenquellen essenziell.

Begründung der Einordnung

Nicht jedes Datenmanagementsystem stellt Datenunabhängigkeit gleichermaßen sicher. Der Grad an Datenunabhängigkeit ist daher eine wichtige Entscheidung, die beim Entwurf eines Systems getroffen werden muss und die zukünftige Nutzung wesentlich beeinflusst. Aus diesem Grund muss der Grad an Datenunabhängigkeit, die ein System bietet, auch bei der Auswahl von Datenmanagementsystemen berücksichtigt werden. Datenunabhängigkeit ist daher klar als Entwurfsprinzip einzustufen.

C.2 Integrität

Entwurfsprinzip

Die Datenintegrität ist ein wichtiges Schutzziel im Rahmen der IT-Sicherheit: „Wir sagen, dass das System die Datenintegrität (engl. Integrity) gewährleistet, wenn es Subjekten nicht möglich ist, die zu schützenden Daten unautorisiert und unbemerkt zu manipulieren“ (Eckert, 2014). Im Datenmanagement spielt die Integrität insbesondere in Zusammenhang mit (oft semantischen) Integritätsbedingungen eine Rolle: Durch diese können einem Datenmanagementsystem deklarativ Bedingungen mitgegeben werden, die der Datenbestand zu jedem Zeitpunkt erfüllen soll. Diese Maßnahme ist dabei eine wichtige Möglichkeit zur Wahrung der Konsistenz eines Datenbestands.

Kernaussagen

- Integrität versucht zu gewährleisten, dass Daten nicht unbemerkt/versehentlich verfälscht werden.
- Die Wahrung der Integrität der gespeicherten Daten ist in den meisten Datenmanagementsystemen zentral.
- Zur Sicherstellung der Integrität können sog. Integritätsbedingungen definiert werden.
- Integritätsbedingungen beschreiben Anforderungen, die der Datenbestand erfüllen soll.
- Ein Datenmanagementsystem das Integrität sicherstellt, ist langsamer als eines, das dies nicht tut.
- Um Daten vor unberechtigter Änderung zu schützen, sind Integritätsbedingungen nicht ausreichend.
- Unbefugte Änderungen können nur durch ausreichenden Zugriffsschutz und Berechtigungsvergabe verhindert werden.
- Daten und Informatiksysteme im Allgemeinen sind angreif- und manipulierbar. Die Reduzierung solcher Möglichkeiten ist in der Informatik zentral.
- Sicherheit besteht nicht nur aus Zugriffsschutz, sondern beinhaltet weitere Aspekte wie Integrität.
- Für die Sicherstellung von Integrität und Sicherheit im Allgemeinen sind weitere Informationen nötig, wie z. B. Integritätsbedingungen, die ein Informatiksystem auswerten kann.

- Nahezu alle Möglichkeiten zur Erhöhung der Sicherheit verlangsamen ein System und/oder beeinträchtigen das Nutzungserlebnis. Der angestrebte Grad an Sicherheit ist daher je nach Anwendungsfall abzuwägen.

Verwandte Konzepte des Datenmanagements

- Konsistenz
- Konkurrenz
- Strukturierung
- Repräsentation

Relevanz in Verbindung zu Praktiken des Datenmanagements

Die Wahrung der Integrität ist immer dann relevant, wenn Daten nicht nur gelesen, sondern auch geschrieben werden können. Dies ist bei den Praktiken Modellierung bis Optimierung der Fall.

Anknüpfungspunkte in der Informatik

- **IT-Sicherheit:** Integrität stellt eines der zentralen Schutzziele der IT-Sicherheit dar.
- **Vergabe von Zugriffsrechten:** Nicht nur in Datenmanagementsystemen ist die Vergabe von Zugriffsrechten zentral, sondern allgemein in Bezug auf Informatiksysteme.
- **Erkennung von Manipulationen:** Die Erkennung von Manipulationen an Datenbeständen, aber auch Datenübertragungen, Software oder Informatiksystemen im Allgemeinen ist in vielen Bereichen der Informatik wichtig.

Begründung der Einordnung

Integritätsbedingungen sind nicht in allen Datenmanagementsystemen gleichermaßen zentral und können – je nach Konzeption eines Systems – auch nicht ohne weiteres eingeführt werden. So ist beispielsweise bei Multimediatelefonbanken oder dokumentenorientierten Datenbanken das Thema „Integrität“ wesentlich komplexer als beispielsweise bei relationalen Datenbanken, bei denen Integritätsbedingungen relativ einfach definiert werden können. Integrität ist daher als Entwurfsprinzip einzustufen.

C.3 Konsistenz

Entwurfsprinzip

In Datenbanken ist Konsistenz seit Jahrzehnten ein zentrales Ziel, das lange Zeit als unumstößlich galt. Konsistenz bezeichnet dabei die Widerspruchsfreiheit des Datenbestands. Diese kann durchgehend oder zu definierten Zeitpunkten erzwungen werden. Im Zusammenhang mit Anfragen an Datenbanken stellt Konsistenz eines der vier im ACID-Prinzip beschriebenen Ziele von Transaktionen dar. Trotz der langjährigen Bedeutung verzichten moderne Datenmanagementsysteme heute teils auf Konsistenz bzw. weichen diese auf (beispielsweise moderne NoSQL-Datenbanken).

Kernaussagen

- Konsistenz liegt vor, wenn ein Datenbestand in sich widerspruchsfrei ist.
- Konsistenz befasst sich mit logischen Widersprüchen in Daten.
- Auftretende Anomalien beim Verändern von Daten müssen verhindert werden, um Konsistenz zu ermöglichen.
- Durch voneinander isolierte Ausführung von Abfragen in Transaktionen kann sichergestellt werden, dass sich nebenläufige Abfragen nicht überschneiden und so Inkonsistenzen herbeiführen.
- Ein durchdachter Aufbau des Datenmodells kann dazu beitragen, Konsistenz zu erreichen, indem Redundanzen vermieden werden.
- Um die Konsistenz von Daten sicherzustellen, müssen über Metadaten oder geeignete Strukturierung zusätzliche Informationen über Daten bereitgestellt werden.
- Das Erkennen von Fehlern in den Daten, zu denen auch Inkonsistenzen zählen, kann nur basierend auf diesen zusätzlichen Informationen stattfinden.
- In vielen Anwendungsfällen ist eine dauerhafte Konsistenz weniger wichtig, sodass eine Aufweichung dieser Anforderung zugunsten der Performanz stattfinden kann.

Verwandte Konzepte des Datenmanagements

- Transaktion
- Integrität
- Konkurrenz
- Redundanz
- Strukturierung
- Synchronisation

Relevanz in Verbindung zu Praktiken des Datenmanagements

Die Wahrung der Konsistenz des Datenbestandes ist bei allen schreibenden Zugriffen, aber auch bei der Entwicklung des Datenmodells, relevant. Dies betrifft daher die Praktiken von Modellierung bis Optimierung.

Anknüpfungspunkte in der Informatik

- **Backup:** Während zur Wahrung der Konsistenz im Allgemeinen dazu tendiert wird, Redundanzen zu vermeiden, stellt ein Backup einen Anwendungsfall dar, in dem eine Inkonsistenz einer redundanten Kopie zu deren Original gezielt angestrebt und genutzt wird.
- **Synchronisation:** Bei der Synchronisation von Daten besteht ein hohes Risiko der Entstehung von Inkonsistenzen. Es müssen Maßnahmen zu deren Vermeidung ergriffen werden.
- **Fehlererkennung:** Die Erkennung von Fehlern funktioniert häufig über die Erkennung von Inkonsistenzen in den Daten bzw. zwischen diesen und über sie gespeicherten Metadaten.

Begründung der Einordnung

Konsistenz war jahrelang eine zentrale Anforderung an Datenmanagementsysteme, stellt heute jedoch eindeutig eine Designentscheidung dar. Je nach Anwendungszweck muss dabei mehr oder weniger Wert auf die konsistente Datenspeicherung gelegt, dafür aber ggf. an anderer Stelle Nachteile in Kauf genommen werden müssen (vgl. CAP-Theorem). Konsistenz ist daher ein Entwurfsprinzip des Datenmanagements.

C.4 Isolierung

Entwurfsprinzip

Die Isolierung zählt zu den zentralen Eigenschaften von transaktionalen Systemen und ist im ACID-Paradigma berücksichtigt. „Diese Eigenschaft verlangt, dass nebenläufig (parallel, gleichzeitig) ausgeführte Transaktionen sich nicht gegenseitig beeinflussen“ (Kemper und Eickler, 2015). Das Prinzip der Isolierung nebenläufiger Transaktionen ist für einen störungsfreien Mehrbenutzerbetrieb in Datenmanagementsystemen unverzichtbar.

Kernaussagen

- Nebenläufige Abfragen können isoliert voneinander durchgeführt werden, obwohl sie sich real überschneiden.
- Durch eine (quasi-)parallele Ausführung von Transaktionen besteht eine hohe Gefahr der gegenseitigen Beeinflussung.
- Gegenseitige Abhängigkeiten der Abfragen können zu Verklemmungen im Datenmanagementsystem führen.
- Lesende Zugriffe sind immer voneinander unabhängig und können daher beliebig parallelisiert werden.
- Eine Isolierung von Abfragen ist nur nötig, wenn mehrere Benutzer gleichzeitig arbeiten.
- Die Isolierung verhindert das Auftreten von Anomalien, die die Konsistenz des Datenbestands gefährden.

Verwandte Konzepte des Datenmanagements

- Transaktion
- Konsistenz
- Konkurrenz

Relevanz in Verbindung zu Praktiken des Datenmanagements

Die Isolierung ist relevant, wenn (mindestens) zwei schreibende Transaktionen auf denselben Datenbestand zugreifen. Sie tritt daher insbesondere in den Praktiken Implementierung und Optimierung zutage.

Anknüpfungspunkte in der Informatik

- **Parallelisierung:** Eine Isolierung von nebenläufigen Aktionen voneinander ist nicht nur im Datenmanagement, sondern beispielsweise auch der (parallelen) Programmierung oder in Zusammenhang mit Betriebssystemen nötig.
- **Mehrbenutzerarchitekturen:** Die Vermeidung von Seiteneffekten, die durch die Nutzung eines Informatiksystems mit mehreren gleichzeitigen Benutzern entstehen, ist in allen Mehrbenutzersystemen zentral.
- **Ablaufsteuerung:** Zur Vermeidung von unerwünschten Seiteneffekten und zur Isolierung von Aktionen voneinander gibt es verschiedene Möglichkeiten der Ablaufsteuerung, wie beispielsweise Sperren.

Begründung der Einordnung

Die Isolierung von Anfragen an ein Datenmanagementsystem muss im Rahmen des Entwurfs und der Implementierung des Systems realisiert werden. Dabei steht auch die Entscheidung an, welche Maßnahmen zur Isolierung umgesetzt werden. Isolierung kann daher als Entwurfsprinzip des Datenmanagements betrachtet werden.

C.5 Dauerhaftigkeit

Entwurfsprinzip

Die Dauerhaftigkeit stellt sicher, dass die Wirkung einer erfolgreich abgeschlossenen Transaktion erhalten bleibt, bis sie durch weitere Änderungen obsolet wird (nach *Kemper und Eickler (2015)*). Dies stellt einen zu jederzeit nachvollziehbaren Datenbestand sicher und erlaubt es Benutzern, sich nach Bestätigung der Transaktionsausführung darauf zu verlassen, dass Änderungen auch korrekt durchgeführt worden sind und nicht, beispielsweise durch andere nicht korrekt isolierte Transaktionen, versehentlich überschrieben werden.

Kernaussagen

- Bestätigte Änderungen des Datenbestands in einem Datenmanagementsystem sollen eine dauerhafte Wirkung zeigen, sodass...
 - bestätigte Änderungen nicht versehentlich verloren gehen.
 - einmal gemachte Änderungen nur durch gezieltes Überschreiben (bzw. durch eine gegenteilige Änderung) rückgängig gemacht werden können.
- Auch im Systemausfall müssen bereits bestätigte Änderungen nachvollzogen und wiederhergestellt werden können.
- Im Falle eines Fehlers innerhalb einer Transaktion dürfen keine Rückstände eines Teils dieser Transaktion im System erkennbar sein.
- Eine übliche Implementierung der Dauerhaftigkeit basiert auf dem Zwei-Phasen-Sperrprotokoll.
- Die Dauerhaftigkeit wird nicht nur durch Nutzung zuverlässiger Speicher, sondern auch durch Replikation und/oder Synchronisation von Daten und durch Protokollierung von Veränderungen sichergestellt.

Verwandte Konzepte des Datenmanagements

- Integrität

- Konsistenz
- Transaktion
- Synchronisation
- Replikation

Relevanz in Verbindung zu Praktiken des Datenmanagements

Da das Prinzip der Dauerhaftigkeit bei schreibendem Zugriff auf das Datenmanagementsystem relevant ist, hat es insbesondere bei Implementierung und Optimierung Bedeutung.

Anknüpfungspunkte in der Informatik

- **Sicherung/Wiederherstellung:** Im Falle eines Systemausfalls müssen Maßnahmen ergriffen werden, um Änderungen beispielsweise aus Protokollen wiederherstellen zu können. Dies ist nicht nur in Datenmanagementsystemen relevant, sondern es stellt sich in der Informatik allgemein die Frage, wie Daten vor Verlust geschützt werden können.
- **Datenspeicherung im Allgemeinen:** Im Zusammenhang mit der Datenspeicherung stellt sich im Allgemeinen die Frage, wie verhindert werden kann, dass durchgeführte Änderungen den Datenbestand gefährden. Um beispielsweise beim Verschieben von Dateien im Dateisystem zu vermeiden, dass bei einem „Verschiebefehler“ die ursprünglichen Dateien verloren sind, kopieren viele Nutzer diese erst und löschen das Original danach (was aufgrund der Implementierung im Betriebssystem eigentlich unnötig ist).

Begründung der Einordnung

Der angestrebte Grad an Dauerhaftigkeit ist eine Entscheidung, die bei der Entwicklung eines Datenmanagementsystems getroffen werden muss. Auch entsprechende Maßnahmen zur Umsetzung müssen in dieser Phase bereits getroffen werden. Somit stellt Dauerhaftigkeit ein Entwurfsprinzip dar.

C.6 Verfügbarkeit

Entwurfsprinzip

Die Verfügbarkeit beschreibt, dass ein Datenmanagementsystem zu einem bestimmten bzw. idealerweise jedem beliebigen Zeitpunkt mit einer angemessenen Reaktionszeit er-

reichbar und nutzbar sein muss (vgl. *Edlich et al. (2011)*). Die Verfügbarkeit von Daten wird in Datenmanagementsystemen im Allgemeinen schon seit langem als zentral angesehen. Im Zusammenhang mit immer häufiger in Echtzeit und von vielen Quellen gleichzeitig erfolgenden Zugriffen wird sie aber immer relevanter.

Kernaussagen

- Bei den meisten Datenmanagementsystemen ist Verfügbarkeit eine zentrale Eigenschaft.
- Verfügbarkeit schließt eine hohe Performanz des Systems ein, aber auch möglichst kurze Ausfallzeiten.
- Verfügbarkeit steht damit in direkter Konkurrenz zu Maßnahmen, die die Sicherheit und Integrität des Datenbestands erhöhen, da diese das System im Allgemeinen verlangsamen und gegebenenfalls zeitweise blockieren.
- Ein besonders hoher Grad an Verfügbarkeit wird in Echtzeitsystemen benötigt, da diese bestimmte Zeitschranken einhalten müssen.
- Verfügbarkeit betrifft nicht nur Datenmanagementsysteme, sondern zeigt sich klar in allen Informatiksystemen und ist besonders auch in sicherheitskritischen Anwendungen zentral.
- Zur Erhöhung der Verfügbarkeit werden Daten oft auf mehrere Knoten repliziert.

Verwandte Konzepte des Datenmanagements

- Integrität
- Konkurrenz
- Replikation
- Transport
- Redundanz

Relevanz in Verbindung zu Praktiken des Datenmanagements

Die Verfügbarkeit ist bei allen Tätigkeiten, bei denen direkt mit dem System bzw. den Daten interagiert wird, zentral. Dies trifft insbesondere auf die Praktiken Implementierung bis Analyse und ggf. auch Visualisierung zu.

Anknüpfungspunkte in der Informatik

- **Informatiksysteme allgemein:** Verfügbarkeit ist bei allen Informatiksystemen ein relevantes Thema. Maßnahmen zur Erhöhung der Verfügbarkeit von Datenmanagementsystemen können daher beispielsweise auch übertragen auf die Verfügbarkeit von Software wiedergefunden werden, indem web- bzw. cloudbasierte Anwendungen oft auf vielen Rechnerknoten repliziert zur Verfügung stehen.
- **Echtzeitsysteme:** Die Verarbeitung von Daten in Echtzeit ist ein immer häufiger anzutreffendes Ziel der Informatik. In allen Fällen spielt Verfügbarkeit eine zentrale Rolle.

Begründung der Einordnung

Verfügbarkeit ist eine Eigenschaft von Datenmanagementsystemen, die in klarem Widerspruch zu anderen Eigenschaften wie Sicherheit und Dauerhaftigkeit steht: Jegliche Maßnahmen zur Erhöhung dieser beiden Eigenschaften senken gleichzeitig die Verfügbarkeit des Systems. Es ist daher je nach Anwendungsfall zu entscheiden, welcher Grad an Verfügbarkeit erreicht werden soll, sodass Verfügbarkeit als Entwurfsprinzip einzuordnen ist.

C.7 Partitionstoleranz

Entwurfsprinzip

Die Partitionstoleranz ist ein Entwurfsprinzip, das in verteilten Datenbanken zum Tragen kommt. Ein Datenmanagementsystem kann, je nach Grad der Partitionstoleranz, den Ausfall der Kommunikation mit einem oder mehreren seiner Knoten verkraften, ohne dass die Funktion eingeschränkt wird oder Fehler bzw. Inkonsistenzen auftreten.

Kernaussagen

- Verteilte Systeme müssen mit dem Ausfall eines oder mehrerer Knoten zurechtkommen.
- Neben dem Ausfall der Knoten ist auch ein Ausfall der Kommunikationsverbindung zu berücksichtigen.
- Bei Ausfällen der Kommunikation kann ein System in mehrere Teile getrennt werden („partitioniert“), die unabhängig voneinander arbeiten.
- Bei Wiedervereinigung des Systems müssen die Änderungen an den Datenbeständen so synchronisiert und nachvollzogen werden, als wäre keine Partitionierung erfolgt.

- Die Partitionstoleranz steht damit aber in klarer Konkurrenz zur Konsistenz und Verfügbarkeit: Hohe Partitionstoleranz bei gleichzeitiger Konsistenz sorgt für Performanceverluste.
- Die Schaffung einer Partitionstoleranz ist nicht nur bei verteilten Datenmanagementsystemen, sondern allgemein bei verteilten Systemen eine zentrale Herausforderung.

Verwandte Konzepte des Datenmanagements

- Konsistenz
- Verfügbarkeit
- Synchronisation
- Transport
- Dauerhaftigkeit
- Transaktion

Relevanz in Verbindung zu Praktiken des Datenmanagements

Partitionstoleranz ist insbesondere im Hintergrund relevant und kommt im Fehlerfall zum Tragen, sodass diese idealerweise nach außen kaum sichtbar wird. Falls doch, dann tritt sie jedoch während allen Interaktionen mit Datenmanagementsystemen, also insbesondere während den Tätigkeiten von Implementierung bis Analyse zutage.

Anknüpfungspunkte in der Informatik

- **Verteilte Systeme:** Der Ausfall von Knoten oder der Kommunikation zwischen diesen ist in allen verteilten Systemen ein relevantes Thema.
- **Rechnerkommunikation:** Im Rahmen der Rechnerkommunikation werden möglichst zuverlässige und sichere Kommunikationsmöglichkeiten entwickelt.

Begründung der Einordnung

Je nach Einsatz und Gestaltung eines Systems kann Partitionstoleranz unverzichtbar aber auch relativ überflüssig sein. Ein gewisser Grad an Partitionstoleranz ist dabei in den meisten verteilten Datenmanagementsystemen zentral. Beim Entwurf und der Implementierung des Systems muss jedoch die Entscheidung getroffen werden, wie stark diese Eigenschaft

priorisiert werden soll, da sie in Konkurrenz zu insbesondere Konsistenz und Verfügbarkeit steht. Damit kann Partitionstoleranz als Entwurfsprinzip angesehen werden.

C.8 Nebenläufigkeit

Entwurfsprinzip

Nebenläufigkeit kann in Zusammenhang mit Datenmanagementsystemen als synonym zur Konkurrenz betrachtet werden, diese betont jedoch nach *Lockemann (1986)* stärker den Wettbewerbsgedanken um Ressourcen. Der Begriff betont damit, dass immer wenn mehrere Nutzer gleichzeitig mit einem Datenmanagementsystem arbeiten, eine Konkurrenzsituation entsteht, mit der das System gewissermaßen umgehen muss.

Der Grad an Nebenläufigkeit, den ein Datenmanagementsystem zulässt, steht dabei in direktem Zusammenhang mit der Menge an Maßnahmen, die dieses ergreifen muss, um andere Prinzipien wie Konsistenz, Integrität, Verfügbarkeit u. Ä. sicherzustellen.

Kernaussagen

- Die verschiedenen Nutzer bzw. Aktionen eines Datenmanagementsystems stehen miteinander in Konkurrenz um die Ressourcen des Systems (insbesondere die Daten).
- Mehrbenutzerbetrieb erhöht zwar i. A. die Verfügbarkeit des Systems, sorgt aber gleichzeitig für Einschränkungen bezüglich anderer Prinzipien: Beispielsweise müssen weitere Maßnahmen ergriffen werden, um die Integrität des Datenbestands aufrechtzuerhalten.
- Ohne explizite Behandlung von Wettbewerbssituationen besteht die Gefahr von Inkonsistenzen und der Entstehung von Fehlern im Datenbestand.
- Konkurrenz kann durch gezielte Ablaufsteuerung in den Griff bekommen werden.
- Zur Auflösung von Konkurrenzsituationen werden Maßnahmen ergriffen wie beispielsweise das Sperren von (Lese- und Schreib-)Zugriffen auf Daten für die Dauer anderer (schreibender) Zugriffe.

Verwandte Konzepte des Datenmanagements

- Transaktion
- Verfügbarkeit
- Integrität
- Konsistenz

Relevanz in Verbindung zu Praktiken des Datenmanagements

Die Konkurrenz bzw. Nebenläufigkeit erlaubt eine gleichzeitige Nutzung der Datenmanagementsysteme durch mehrere Nutzer. Daher spielt dieses Entwurfsprinzip bei allen Tätigkeiten, in denen direkt Datenmanagementsysteme verwendet werden, d. h. insbesondere in den Phasen Implementierung bis Analyse eine große Rolle.

Anknüpfungspunkte in der Informatik

- **Mehrbenutzersysteme im Allgemeinen:** Nebenläufigkeit spielt in allen Mehrbenutzersystemen eine zentrale Rolle und muss jeweils geeignet beherrscht werden.
- **Parallelisierung:** Auch im Rahmen von Parallelisierung spielen dieselben Konzepte und Maßnahmen wie bei Nebenläufigkeit bzw. Konkurrenz eine wichtige Rolle.
- **Betriebssysteme:** In Betriebssystemen sind die Konzepte der Konkurrenz bzw. Nebenläufigkeit besonders zentral, da diese sich hier nicht nur bei Mehrbenutzerbetrieb, sondern auch im Bereich des Multi-Tasking auswirken.

Begründung der Einordnung

Der Grad an Nebenläufigkeit, den ein System zulässt, ist eine Entscheidung, die im Rahmen der Entwicklung eines Datenmanagementsystems getroffen und die gegenüber anderen Entwurfsprinzipien abgewogen werden muss. Es handelt sich daher auch bei Nebenläufigkeit um ein Entwurfsprinzip.

C.9 Redundanz

Entwurfsprinzip

Redundanz tritt auf, wenn identische Daten als Kopien an verschiedenen Orten gespeichert werden. Da Redundanzen die Gefahr bergen, zu Anomalien und damit Inkonsistenzen im Datenbestand zu führen, gilt es häufig, diese zu vermeiden. Gleichzeitig wird Redundanz jedoch auch an verschiedenen Stellen gezielt eingesetzt, beispielsweise um Effizienzsteigerungen zu bewirken, aber auch zur Erhöhung der Ausfallsicherheit und Verfügbarkeit von Systemen.

Kernaussagen

- Redundante Datenspeicherung birgt das Risiko der Entstehung von Inkonsistenzen im Datenbestand.

- Zur Vermeidung von Redundanz können Datenbestände beispielsweise normalisiert werden.
- Redundante Daten können zur Erhöhung der Verfügbarkeit eines Systems im Sinne der Replikation der Daten eingesetzt werden.
- Durch Redundanz kann die Sicherheit von Daten erhöht werden, da ggf. verlorene Daten wiederhergestellt werden können.
- Redundanz kann auf verschiedener konzeptioneller Ebene eingesetzt werden bzw. auftreten: Beispielsweise können Informationen in einer Kopie eines Datensatzes redundant sein, aber auch die physischen Daten redundant gespeichert werden.

Verwandte Konzepte des Datenmanagements

- Integrität
- Konsistenz
- Verfügbarkeit
- Replikation
- Synchronisation
- Strukturierung

Relevanz in Verbindung zu Praktiken des Datenmanagements

Die Vermeidung oder Zulassung von Redundanz ist insbesondere in den Planungsphasen des Datenmanagements zentral. Dies betrifft daher die Praktiken Modellierung bis Optimierung.

Anknüpfungspunkte in der Informatik

- **Dateispeicherung allgemein:** Selbst bei Verwendung normaler Dateisysteme kann die Vermeidung von Redundanz zielführend sein, indem beispielsweise Links genutzt werden, anstatt Dateien in Kopie abzuspeichern, da auch dabei Änderungsanomalien auftreten können.
- **Backup:** Backups stellen eines der zentralen Beispiele für die Nutzung von Redundanz dar.
- **Programmierung:** Die hinter dem Prinzip der Redundanznutzung und -vermeidung liegende Idee kann auch in der Programmierung angetroffen werden, die im Allgemeinen eher versucht, Dopplungen im Programmcode aus demselben Grund zu

vermeiden wie bei Datenmanagementsystemen. Auch hier kann es jedoch stellenweise sinnvoll sein, Redundanz auszunutzen um Code verständlicher zu machen, ein Nachladen von Modulen zu verhindern o. Ä.

- **Fehlertoleranz:** Auch abseits des Datenmanagements kann Redundanz zur Erhöhung der Fehlertoleranz eines Gesamtsystems eingesetzt werden, indem beispielsweise redundante Server vorgehalten werden, die bei Ausfall eines Servers automatisiert einspringen können.
- **Fehlererkennung:** Auch zur Erkennung von Fehlern eines Computersystems kann Redundanz hilfreich sein, indem beispielsweise Berechnungen in sicherheitsrelevanten Bereichen (bspw. Fahrstraßensteuerung bei Bahnstellwerken) durch drei unabhängig arbeitende redundante Computer durchgeführt werden und nur als korrekt angenommen werden, falls mindestens zwei dieser Rechner zum selben Ergebnis kommen.

Begründung der Einordnung

Die Nutzung und/oder Vermeidung von Redundanz auf unterschiedlichen Ebenen zeigt, dass es sich bei diesem Prinzip um eine Entscheidung handelt, die im Rahmen des Entwicklungsprozesses eines Datenmanagementsystems, aber auch bei der Strukturierung von Daten getroffen werden muss. Redundanz stellt daher klar ein Entwurfsprinzip des Datenmanagements dar.

C.10 Strukturierung

Mechanismus

Strukturierung ist einer der zentralen Aspekte bei der Speicherung und Verwaltung von, aber auch beim Zugriff auf Daten: Diese müssen immer strukturiert abgelegt werden, damit ein gezielter Zugriff auf diese möglich ist; die Struktur von gespeicherten Daten wird dabei stark durch die Art der Daten und den Einsatzzweck beeinflusst und kann drastische Auswirkungen auf Performanz und Nutzbarkeit von Systemen haben. Daher können geeignete Primär- und Sekundärstrukturen auch eingesetzt werden, um ein System in diesem Zusammenhang zu optimieren.

Kernaussagen

- Um einen Zugriff auf Daten zu ermöglichen, muss immer eine gewisse Strukturierung vorgenommen werden.
- Für die Vermeidung von Inkonsistenzen bietet sich eine möglichst starke Strukturierung der Daten an.

- Eine starke Strukturierung sorgt meist für eine starke Partitionierung des Datenbestands und senkt damit die Performanz des Zugriffs.
- Zur Strukturierung werden in vielen Fällen Metadaten eingesetzt.
- Neben der Strukturierung der eigentlichen Daten durch Festlegung von Primärstrukturen, können zusätzliche Sekundärstrukturen festgelegt werden, die alternative Wege zum Zugriff auf Daten eröffnen.
- Die Strukturierung von Daten kann auf verschiedenen Ebenen des Datenmanagementsystems erfolgen, der physikalischen, logischen oder konzeptuellen Ebene.
- Das im Rahmen der Datenmodellierung meist genutzte und auch am nächsten am Nutzer gelegene Modell ist das konzeptionelle Modell.

Verwandte Konzepte des Datenmanagements

- Datenunabhängigkeit
- Integrität
- Konsistenz
- Redundanz
- Repräsentation

Relevanz in Verbindung zu Praktiken des Datenmanagements

Die Strukturierung ist, je nach Ebene, auf der sie stattfindet, insbesondere in Zusammenhang mit der Modellierung, Implementierung und Optimierung von Bedeutung.

Anknüpfungspunkte in der Informatik

- **Rechnerkommunikation:** Strukturierende Ansätze finden sich in der Rechnerkommunikation beispielsweise beim Aufbau von Rechnernetzen und der Strukturierung von Kommunikation zwischen verschiedenen Systemen.
- **Softwareengineering:** Im Softwareengineering ist die Strukturierung sowohl für den Projektverlauf als auch die Planung des konkreten Produkts zentral.
- **Theoretische Informatik:** In der theoretischen Informatik werden strukturierende Aspekte genutzt, um Probleme zu modellieren, aber auch um Zusammenhänge verschiedener Konzepte der theoretischen Informatik zu betonen (z. B. Chomsky-Hierarchie).

Begründung der Einordnung

Obwohl der Grad an Strukturierung je nach System, Verwendungszweck und Daten unterschieden werden muss, handelt es sich bei der Strukturierung nicht um ein Entwurfsprinzip: Strukturierung wird im Datenmanagement insbesondere eingesetzt, um verschiedene Entwurfsprinzipien zu erreichen und nicht zum Selbstzweck. Es handelt sich bei der Strukturierung daher eher um ein Prinzip, das die korrekte Funktion eines Datenmanagementsystems ermöglicht, d. h. um einen Mechanismus.

C.11 Repräsentation

Mechanismus

Unter Repräsentation werden alle Aspekte der internen Speicherung von Informationen in einem Datenmanagementsystem verstanden. Dies beinhaltet Themen wie Datenstrukturen (z. B. Suchbäume), Optimierungen in diesem Bereich (z. B. Kompression, Einsatz von Pufferspeichern) und grundsätzliche Möglichkeiten zur Datenspeicherung (bspw. die Nutzung von Vorder- und Hintergrundspeichern).

Kernaussagen

- Die Repräsentation erfolgt auf unterschiedlichen Ebenen:
 - Informationen müssen zur Speicherung im Datenmanagementsystem als Daten repräsentiert werden.
 - Die Daten werden wiederum für die physikalische Speicherung geeignet repräsentiert.
 - Zur Organisation der Daten werden diese geeigneten Datenstrukturen gespeichert, die sich je nach Art der Daten und Ziel des Datenmanagementsystems unterscheiden.
- Bei der Speicherung von Daten muss zwischen Vorder- und Hintergrundspeichern unterschieden werden.
- Zur Optimierung der Performanz eines Systems können Daten zusätzlich in einem Pufferspeicher vorgehalten werden.
- Das Datenvolumen kann durch Kompression in gewissen Grenzen verringert werden.
- Ohne geeignete Interpretationsvorschriften sind die gespeicherten Daten oft wertlos, da die Information nicht zurückgewonnen werden kann.

Verwandte Konzepte des Datenmanagements

- Strukturierung
- Dauerhaftigkeit
- Transport

Relevanz in Verbindung zu Praktiken des Datenmanagements

Die geeignete Repräsentation von Daten findet üblicherweise im Hintergrund statt und tritt kaum bei den Praktiken des Datenmanagements zutage, außer beispielsweise durch Auswahl von Datentypen, die über die Repräsentation mitentscheiden. Sie ist damit höchstens im Bereich der Modellierung und Implementierung erkennbar. Die zweite Ebene der Repräsentation, nämlich die Darstellung der Informationen als Daten, ist hingegen klar bei der Datenerfassung/-gewinnung relevant, aber auch der Datenbereinigung.

Anknüpfungspunkte in der Informatik

- **Theoretische Informatik:** Die Repräsentation von Informationen wird in der theoretischen Informatik beispielsweise durch die Betrachtung des Informationsgehalts thematisiert.
- **Datenstrukturen:** Die im Bereich der Datenstrukturen thematisierten Möglichkeiten zur Speicherung von Daten finden im Datenmanagement starke Anwendung zur Repräsentation der Daten.
- **Rechnerkommunikation:** In der Rechnerkommunikation müssen Daten so geeignet repräsentiert werden, dass sie zuverlässig und fehlerfrei übertragen werden können.

Begründung der Einordnung

Die Repräsentation findet eher in den ersten Schritten des Datenmanagements, bei der Entwicklung der Systeme und der Gewinnung der Daten, statt. Durch sie werden alle weiteren Möglichkeiten des Datenmanagements erst umsetzbar. Es handelt sich daher bei der Repräsentation um einen dem Datenmanagement zugrundeliegenden Mechanismus.

C.12 Replikation

Mechanismus

Replikation bezeichnet die gezielte redundante Speicherung von Daten an mehreren Orten / auf mehreren Datenspeichern. In verteilten Systemen werden Daten möglichst sofort repliziert, sodass zu jedem Zeitpunkt eine Kopie der aktuellen Daten vorliegt, die beispielsweise im Falle eines Ausfalls eines Datenspeichers als Ersatz genutzt werden kann. Je nach Ziel kann die Replikation synchron ablaufen, d. h. die eigentliche Abfrage solange blockieren, bis die Replikation abgeschlossen ist, oder asynchron und ohne Blockierung im Hintergrund stattfinden. Auch im Heimanwenderbereich bekommt Replikation – beispielsweise in Zusammenhang mit Netzwerkspeichersystemen (NAS) – zunehmende Bedeutung: Hier wird Replikation eingesetzt, indem Festplatten in RAID-Arrays organisiert werden, die Daten u. A. auf zwei Festplatten repliziert speichern können.

Kernaussagen

- Daten können – automatisiert oder manuell – auf verschiedene Datenspeicher repliziert werden.
- Durch die Replikation kann sowohl die Ausfallsicherheit eines Informatiksystems im Gesamten, aber auch die Verlustsicherheit der Daten selbst erhöht werden.
- Gleichzeitig sorgt Replikation durch die dadurch erzeugte Redundanz auch für potenzielle Einschränkungen der Konsistenz der Daten.
- Je nach angestrebter Form der Replikation – d. h. synchron oder asynchron – kann diese die Performanz des Datenmanagementsystems negativ beeinflussen.
- Während eine asynchrone Replikation zwar positive Einflüsse auf die Performanz des Systems hat, wird dadurch die Dauerhaftigkeit und Konsistenz eingeschränkt, da nicht garantiert werden kann, dass im Fehlerfall bereits alle bestätigten Änderungen auch repliziert wurden.

Verwandte Konzepte des Datenmanagements

- Konsistenz
- Integrität
- Verfügbarkeit
- Dauerhaftigkeit
- Partitionstoleranz
- Redundanz

- Synchronisation
- Transport
- Transaktion

Relevanz in Verbindung zu Praktiken des Datenmanagements

Replikation findet im Datenmanagementsystem im Hintergrund statt und sollte durch den Nutzer kaum bemerkbar sein. Daher muss sie insbesondere bei der Implementierung und Optimierung beachtet werden, aber auch bei der Löschung von Daten, die redundante Kopien miteinbeziehen muss.

Anknüpfungspunkte in der Informatik

- **Verteilte Systeme:** Die Replikation spielt in verteilten Systemen allgemein eine Rolle, da, selbst wenn das System kein Datenmanagementsystem darstellt, gewisse Daten repliziert werden müssen. Gleichzeitig wird das Grundprinzip der Replikation, die redundante Vorhaltung zueinander äquivalenter Ressourcen, auch ansonsten vielfältig bei solchen Systemen eingesetzt.
- **Rechnerkommunikation:** Jegliche Verfahren zur Replikation basieren auf solchen aus dem Fachgebiet Rechnerkommunikation. Bei der Replikation spielen verschiedene Konzepte aus der Rechnerkommunikation, wie die synchrone oder asynchrone Datenübertragung, die Grenzen dieser Kommunikation, etc. eine deutliche Rolle.
- **Cloud-Computing:** Im immer wichtiger werdenden Bereich des Cloud-Computing ist Replikation heute zentral, da nur durch Replikation ein Cloud-System ausfallsicher und performant entworfen werden kann.

Begründung der Einordnung

Als eher technischer Aspekt, der in Datenmanagementsystemen im Hintergrund relevant ist, wurde die Replikation als Mechanismus eingeordnet, der im offensichtlicheren Bereich zur Erreichung verschiedener Entwurfsprinzipien beiträgt.

C.13 Synchronisation

Mechanismus

Im Bereich des Datenmanagements können zwei Arten der Synchronisation erkannt werden: Einerseits die Synchronisation von Daten, die über reine Replikation in der Hinsicht

hinausgeht, dass Entscheidungen darüber getroffen werden, wie beispielsweise mit konkurrierenden Änderungen umgegangen wird. Andererseits bezeichnet Synchronisation auch die Koordination gleichzeitiger bzw. konkurrierender Zugriffe auf Datenbestände, die immer dann erfolgen muss, wenn konkurrierende Zugriffe auf Daten erlaubt werden sollen, die sich jedoch nicht beeinflussen dürfen und voneinander unbemerkt ablaufen müssen.

Kernaussagen

- Bei der Synchronisation von Daten entstehen Konflikte, die aufgelöst werden müssen.
- Eine automatische Lösung von Synchronisationskonflikten ist oft nicht möglich.
- Die Synchronisation von Daten kann (theoretisch) auf beliebig viele Geräte/Speicher stattfinden.
- Zur Vermeidung einer gegenseitigen Beeinflussung von konkurrierenden Abfragen bzw. Zugriffen müssen diese geeignet synchronisiert werden.
- Zur Synchronisation werden verschiedene Techniken eingesetzt, wie beispielsweise Sperren oder die Synchronisierung anhand von Zeitstempeln.
- Synchronisierungsmaßnahmen werden insbesondere eingesetzt, um dafür zu sorgen, dass eine serialisierbare Abfolge an Abfragen entsteht.
- Ein verbreitetes Sperrprotokoll stellt das Zwei-Phasen-Sperrprotokoll dar.

Verwandte Konzepte des Datenmanagements

- Integrität
- Isolierung
- Transaktion
- Konkurrenz
- Replikation
- Transport

Relevanz in Verbindung zu Praktiken des Datenmanagements

Die Synchronisation wird in denselben Bereichen relevant, in denen auch Konkurrenz auftreten kann. Es handelt sich daher um alle Praktiken, bei denen direkt Datenmanagementsysteme genutzt werden, d. h. insbesondere in den Phasen Implementierung bis Analyse.

Anknüpfungspunkte in der Informatik

- **Betriebssysteme:** Die Synchronisation spielt in Betriebssystemen insbesondere auf Prozessebene eine zentrale Rolle und wird mit ähnlichen Maßnahmen wie im Datenmanagement gelöst.
- **Parallele Programmierung:** In der parallelen Programmierung ist die Synchronisation zentral, um beispielsweise Zugriffe auf gemeinsam genutzte Variablen (bzw. Ressourcen allgemein) zu kontrollieren.

Begründung der Einordnung

Synchronisation ist eine technische Lösung, die es ermöglicht, Datenmanagementsysteme mit verschiedenen Nutzern nebenläufig zu verwenden und/oder Daten lokal vorzuhalten, ohne mit einem System kontinuierlich kommunizieren zu müssen. Sie kann daher klar als Mechanismus des Datenmanagements eingeordnet werden.

C.14 Partitionierung

Mechanismus

Ähnlich wie bei der Replikation findet bei der Partitionierung von Daten eine Nutzung mehrerer Datenspeicher zur Speicherung statt. Im Gegensatz zur Replikation werden jedoch nicht dieselben Daten auf mehreren Datenspeichern gespeichert, sondern stattdessen die Daten auf mehrere Speicher (nicht-redundant) verteilt. Dies ist daher insbesondere hilfreich, wenn die Kapazität eines einzelnen Datenspeichers nicht mehr ausreicht oder zur Erhöhung der Performanz des Gesamtsystems auf mehrere Datenspeicher gleichzeitig geschrieben werden soll.

Kernaussagen

- Daten können durch Partitionierung auf verschiedene Datenspeicher verteilt werden.
- Partitionierung kann die Gesamtkapazität eines Systems über die Kapazität eines einzelnen Speichervolumens hinaus erhöhen.
- Durch Partitionierung kann die Geschwindigkeit des Gesamtsystems erhöht werden, indem auf mehrere Datenspeicher parallel geschrieben bzw. von diesen gelesen wird.
- Partitionierung kann für erhöhte Anfragezeiten sorgen, wenn einzelne Datenknoten ausfallen oder langsam reagieren, obwohl das Gesamtsystem verfügbar scheint.
- Bei Partitionierung der Daten werden Maßnahmen zur Sicherung der Integrität des Datenbestandes aufwendiger.

Verwandte Konzepte des Datenmanagements

- Partitionstoleranz
- Integrität
- Verfügbarkeit
- Transport

Relevanz in Verbindung zu Praktiken des Datenmanagements

Da die Partitionierung eher im Hintergrund relevant ist, sollte sie höchstens im Bereich der Implementierung und Optimierung relevant werden, da sie die Nutzbarkeit des Gesamtsystems während dieser Praktiken beeinflusst. Idealerweise tritt sie aber für den Nutzer nicht augenscheinlich zutage.

Anknüpfungspunkte in der Informatik

- **Parallelisierung:** Im Rahmen der Parallelisierung von Aufgaben werden diese in Teilaufgaben (gleicher oder unterschiedlicher Art) zerlegt, die auf mehreren Rechenknoten parallel abgearbeitet werden können.
- **Blockchain:** Bei Blockchain-Algorithmen werden die Informationen über die gesamte Kette dezentral vorgehalten, indem viele Datenblöcke kryptografisch miteinander verknüpft werden.

Begründung der Einordnung

Partitionierung ist eine Maßnahme, die im Hintergrund im Datenmanagementsystem erfolgt und hauptsächlich aus technischen Gründen getroffen wird, da Beschränkungen einzelner Datenspeicher hinsichtlich ihrer Kapazität oder Performanz aufgebrochen werden müssen. Es handelt sich damit bei der Partitionierung klar um einen Mechanismus des Datenmanagements.

C.15 Transport

Mechanismus

Beim Transport handelt es sich um die Übertragung von Daten innerhalb eines Systems oder über dessen Grenzen hinweg. Die Übertragung der Daten muss dabei – je nach Zweck – verschiedenen Kriterien genügen und beispielsweise eine hohe Vertraulichkeit, Geschwin-

digkeit oder Fehlertoleranz sicherstellen. Obwohl die Funktionsweise des Transports eher Thema des Fachgebiets Rechnerkommunikation ist, stellt dieser einen wichtigen Mechanismus für das Datenmanagement dar, da Datenmanagementsysteme erst dadurch sinnvoll nutzbar werden. Gleichzeitig müssen Eigenschaften des Transports von Daten und der Fakt, dass dieser geschieht, auch bei der Implementierung von Datenmanagementsystemen beachtet werden, um dadurch entstehende Fehler und Verzögerungen zu berücksichtigen.

Kernaussagen

- Der Transport erfolgt mit üblichen Protokollen und mittels üblicher Methoden der Rechnerkommunikation.
- Je nach Art und Zweck des Systems kann es sinnvoll sein, dieses möglichst verteilt zu organisieren, sodass sehr viel Transport nötig wird, aber auch es möglichst zentralisiert aufzubauen und Transport soweit möglich zu vermeiden.
- Umso mehr Transport von Daten nötig ist, umso langsamer wird in der Regel das Datenmanagementsystem.
- Daten müssen für den Transport in der Regel verschlüsselt und mit zusätzlichen Informationen, wie Metadaten zur Absicherung des Transports und zur Erhöhung der Fehlertoleranz, versehen werden.
- Die Sicherstellung von Konsistenz sorgt in verteilten Systemen für besonders viel Kommunikation.

Verwandte Konzepte des Datenmanagements

- Partitionierung
- Synchronisation
- Replikation
- Verfügbarkeit
- Integrität

Relevanz in Verbindung zu Praktiken des Datenmanagements

Transport wird in Zusammenhang mit verschiedenen Praktiken relevant, immer dann, wenn Daten mit dem System, innerhalb des Systems oder vom Nutzer mit anderen ausgetauscht werden. Dies ist insbesondere bei der Datenerfassung/-gewinnung, der Implementierung, der Analyse und dem Austausch der Fall, wobei auch bei vielen der anderen Praktiken in Ansätzen Zusammenhänge zu diesem Mechanismus existieren.

Anknüpfungspunkte in der Informatik

- **Programmierung:** In der Programmierung tritt Transport insbesondere in Form von Kommunikation zwischen Modulen auf. Umso höher die Kopplung zweier Module zueinander ist, umso mehr Kommunikation wird für deren Aufgaben nötig.
- **Rechnerkommunikation:** Die Organisation des Transports von Daten ist die zentrale Aufgabe des Fachgebiets Rechnerkommunikation.

Begründung der Einordnung

Obwohl es sich beim Transport prinzipiell um ein Thema des Fachgebiets Rechnerkommunikation handelt, ist dieses für die Realisierung von Datenmanagementsystemen zentral. Ohne den Transport von Daten wäre die Funktion von diesen Systemen undenkbar – somit stellt Transport auch einen Mechanismus des Datenmanagements dar.

C.16 Transaktion

Mechanismus

Transaktionen werden in Datenmanagementsystemen als Mechanismus genutzt, um einzelne Abfragen zu einer Gruppe von Abfragen zu gliedern, die nur gemeinsam ausgeführt werden. Auf diese Weise kann sichergestellt werden, dass eine Änderung, die aus mehreren Teilabfragen besteht, nur entweder komplett durchgeführt wird oder komplett ohne Wirkung verbleibt. Typischerweise erfolgen Maßnahmen zur Beherrschung von Konkurrenz insbesondere auf Transaktionsebene, sodass Transaktionen bevor sie komplett ausgeführt worden sind keinerlei Einfluss auf andere Transaktionen bzw. Abfragen haben und somit Abfragen unabhängig voneinander scheinen.

Kernaussagen

- Transaktionen werden genutzt, um eine dauerhafte Konsistenz des Datenbestands sicherzustellen.
- Transaktionen werden nur komplett oder gar nicht ausgeführt.
- Im Fehlerfall während der Ausführung einer Transaktion wird diese komplett rückgängig gemacht („rollback“).
- Die Koordination konkurrierender Zugriffe erfolgt typischerweise auf Transaktionsebene.

- Die nicht-Atomarität von Transaktionen ermöglicht es, dass zwei nebenläufige Transaktionen gegenseitig voneinander so abhängen, dass dieser Konflikt nicht ohne Weiteres aufgelöst werden kann („deadlock“). In einem solchen Fall wird meist eine Transaktion abgebrochen.

Verwandte Konzepte des Datenmanagements

- Konsistenz
- Isolation
- Dauerhaftigkeit
- Konkurrenz

Relevanz in Verbindung zu Praktiken des Datenmanagements

Die Transaktionalität eines Datenmanagementsystems ist bei der schreibenden Interaktion mit dem System relevant, obwohl sie idealerweise eigentlich für den Nutzer unbemerkbar sein sollte. Damit ist sie insbesondere in Zusammenhang mit den Praktiken Implementierung und Optimierung zentral.

Anknüpfungspunkte in der Informatik

- **Parallelisierung:** Auch bei der Parallelisierung anderer Aufgaben tritt ein den Transaktionen ähnliches Konzept zutage: Beispielsweise gibt es bei der parallelen Programmierung kritische Abschnitte, deren Anweisungen nicht unterbrochen werden dürfen, d. h. dass diese, wie Transaktionen, nur gemeinsam oder gar nicht ausgeführt werden.
- **Rechnerkommunikation:** In der Rechnerkommunikation tritt ein ähnliches Konzept auf, wenn nach einer Anfrage an einen anderen Rechner nur Teile der Antwort übermittelt werden können, da dann entweder versucht werden muss, diese Teile erneut anzufordern oder die Anfrage im Gesamten als fehlerhaft verworfen und neu ausgeführt werden muss.

Begründung der Einordnung

Transaktionen stellen ein Konzept von Datenmanagementsystemen dar, das im Hintergrund von diesen Systemen verwendet wird und auf technischer Ebene die Erfüllung verschiedener Entwurfsprinzipien ermöglicht. Es ist daher für die Funktionsweise von Datenmanagementsystemen aus technischer Sicht zentral und kann somit als Mechanismus des Datenmanagements betrachtet werden.