

1 Einführung in die Welt der Datierung in die Welt der Daten

Sonntag, 10. September 2023 15:56



Daten systemübergreifend bereitstellen *Einführung in die Welt der Daten*

11. Klasse
AWP

1. Einführung in die Welt der Daten

Daten sind das Herz der IT. Sei es beim maschinellen Arbeiten in einem Computer und bei der Kommunikation zwischen Computern, immer fallen Daten in den unterschiedlichsten Datenformaten an. Gleichmaßen besteht das menschliche Leben aus Daten. Geburtsdatum, Adressdaten, Bewegungsdaten und so weiter ...



Arbeitsauftrag

- Lesen** Sie den nachfolgenden Beitrag von Michael Kroker.
- Bringen** Sie den Ausspruch „DATEN SIND DAS ÖL DES 21. JAHRHUNDERTS“ in den **Kontext** des Beitrags.



Zwei Drittel der Unternehmen erwarten, dass sich die Datenmengen bis 2025 verfünffachen

31. August 2020, Autor: **Michael Kroker**

Gleichwohl sind fast drei Viertel der Unternehmen in Deutschland überzeugt davon, dass sie die Flut von Daten erfolgreich meistern werden.



Internet of Things und die umfassende Vernetzung aller Maschinen, Gadgets und Geräte über das Internet sorgen für exponentielles Datenwachstum. Das wiederum bringt Chancen und Herausforderungen für die Unternehmen mit sich.

Die überwiegende Mehrheit der Unternehmen hält Daten für äußerst oder sehr wertvoll – gleichzeitig befürchtet jedoch eine Mehrheit von 57 Prozent, die Datenmenge werde so schnell anwachsen, dass ihr Unternehmen nicht Schritt halten kann. Zwei Drittel der Unternehmen weltweit erwarten, dass sich allein die Datenmenge bis 2025 nahezu verfünffachen wird (67 Prozent).

Das sind die zentralen Ergebnisse einer Umfrage des Marktforschungsunternehmens True Global Intelligence im Auftrag des amerikanischen Software-Anbieters Splunk unter weltweit 2259 Führungskräften aus Business und IT, darunter Deutschland, Großbritannien, Frankreich, Niederlande, sowie USA, China und Australien. „Kroker's Look @ IT“ veröffentlicht die Ergebnisse der Studie heute exklusiv vorab.

Führungskräfte erkennen die große Chance dieses explosiven Datenwachstums und halten Daten für äußerst oder sogar sehr wertvoll für den Gesamterfolg (81 Prozent), die Innovationskraft (75 Prozent) und die Cybersicherheit (78 Prozent) ihrer Unternehmen.

Dennoch berichten 66 Prozent der Befragten weltweit, dass mindestens die Hälfte ihrer Unternehmensdaten Dark Data seien, also unerschlossene, ungenutzte oder völlig unbekannte Daten. Das ist ein 10-prozentiger Anstieg im Vergleich zum Vorjahr.

Die Studienergebnisse aus Deutschland zeigen, dass Unternehmen hierzulande im Jahr 2025 mit einem 4,5-fachen Anstieg der Datenmenge im Vergleich zu heute rechnen. Über die Hälfte der 250 Befragten (44 Prozent Mittelstand, 56 Prozent Großunternehmen) gibt an, dass das Datenvolumen bereits jetzt so schnell ansteigt, dass ihr Unternehmen nicht mehr mit der Entwicklung Schritt halten kann. Die Unternehmen spüren die Datenwelle, bereiten sich jedoch schon aktiv darauf vor oder sind bereits vorbereitet.

Weitere Ergebnisse aus Deutschland:

- Positive Prognose: Neun von zehn deutschen Befragten erklären, dass ihre Unternehmen sich bereits gezielt zukunftssicher aufstellen, um für die kommende Datenwelle gewappnet zu sein (86 Prozent). Nur jeder fünfte deutsche Befragte gibt an, dass es in ihrem Unternehmen eine große Verunsicherung diesbezüglich gibt (19 Prozent). Vier von zehn verspüren hingegen keinerlei oder nur eine geringe Verunsicherung (41 Prozent). Es zeigt sich eine generell positive Grundeinstellung: Fast drei Viertel der Unternehmen sind überzeugt davon, dass sie die Flut von Daten erfolgreich meistern werden. Nur die Befragten aus China sind mit 96 Prozent noch positiver eingestellt.
- Dark Data als Herausforderung und Chance zugleich: Mehr als jeder zweite deutsche Befragte schätzt, dass es sich bei mindestens der Hälfte ihrer Unternehmensdaten um Dark Data handelt (55 Prozent). Die Mehrheit der Befragten sieht die Gründe dafür in der Technologie oder den technischen Prozessen – beides Gründe, die durch ein optimiertes Datenmanagement und neue Softwarelösungen behoben werden könnten. Darüber hinaus begreifen 71 Prozent der Befragten die Datenwelle auch als Chance, sich proaktiv mit Dark Data in ihrem Unternehmen auseinanderzusetzen. 62 Prozent der deutschen Unternehmen gehen dieses Problem bereits an: Sie versuchen, der Datenwelle einen Schritt voraus zu sein, indem sie Dark Data identifizieren und nutzbar machen.
- Hindernisse für wichtige Technologien: Die sechs wichtigen Zukunftstechnologien KI, 5G, Blockchain, IoT, AR/VR, Edge Computing stoßen in Deutschland hauptsächlich auf strukturelle Hindernisse. Am häufigsten werden folgende genannt:
 - Mangel an Fachpersonal
 - Schwierigkeiten beim Identifizieren von Dark Data
 - Schwierigkeiten bei der Verwaltung und Nutzung von Daten
 - Mangelndes Technologieverständnis

1.1 Was sind Daten und was sind Informationen?



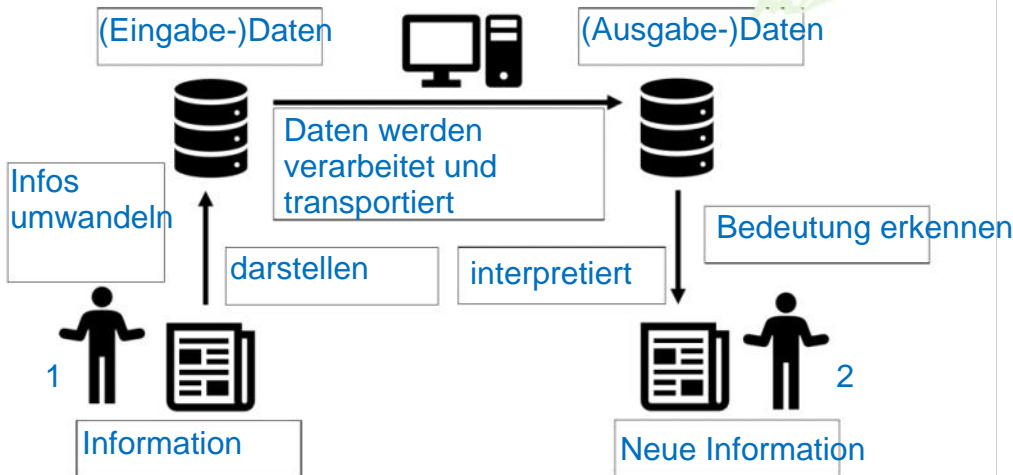
Der Begriff „Information“ wird umgangssprachlich häufig für Daten verwendet. Die Informatik unterscheidet dagegen zwischen Daten und Information. Zur Übertragung oder Verarbeitung von Nachrichten bildet der Mensch die Information in Daten ab. Daten sind maschinell verarbeitbare Zeichen. Die in einer Nachricht enthaltene Information stellt die Bedeutung der Nachricht dar. Die Übertragung oder Verarbeitung geschieht dann auf der Ebene der Daten, und das Ergebnis wird wiederum vom Menschen als Information interpretiert.

Die Repräsentation von Information durch Daten und die Interpretation der Daten als Information ist eine gedankliche Leistung des Menschen und findet nicht im datenverarbeitenden System statt. Informatisch Geschulte nutzen geeignete Modelle, um Information durch Daten zu repräsentieren. Solche geeigneten Modelle beinhalten zugleich Darstellungsformen, bei der einerseits die gewünschte automatische Verarbeitung möglich ist und andererseits die Ergebnisse von Menschen wiederum als Information interpretiert werden können.



Arbeitsauftrag

Ergänzen Sie anhand des obenstehenden Textes die Skizze



1.2 Übersicht verschiedener strukturierter Datenformate



Zur Datenhaltung und zum Datenaustausch existieren viele Datenformate. Der Mensch kann aus Texten, Bildern usw. verschieden Informationen extrahieren. Für den Computer sind diese Texte aber unstrukturiert und ohne weitere Zusätze wie sogenannte Marker nicht zu „verstehen“. Erst die Auszeichnung bzw. Markierung und der in der Regel damit verbundene logische und meist auch hierarchische Aufbau eines Dokuments sorgen für die „Verständlichkeit“. Strukturierte Daten werden dabei in einem bestimmten, definierten Format strukturiert. Es gibt verschiedene Formate für die Strukturierung von Daten. Nicht jede Software kann jedes strukturierte Datenformat lesen und bearbeiten.¹



Arbeitsauftrag

Alles braucht eine Struktur!



- Informieren** Sie sich mittels Recherche über die Datenformate **JSON, CSV, XML & YAML**.
- Füllen Sie die nachfolgende Tabelle aus.**
- Kreuzen** Sie das Format an, welches Sie schon einmal in Ihrer **betrieblichen Arbeit** nutzt.
- Bereiten Sie sich auf die **Atom-Molekül-Diskussion** vor.
 - Suchen Sie sich einen **Partner*In** und diskutieren Sie **5 Minuten** über die Aussage: „XML IST DAS BESTE DATENFORMAT UND SOLLTE STETS VERWENDET WERDEN!“
 - Suchen Sie sich im Anschluss ein **weiteres Pärchen** und **diskutieren** Sie nun weitere **5 Minuten** zu viert über die Aussage.
 - Im Anschluss sucht sich **jede 4er-Gruppe eine weitere 4er-Gruppe** und spricht **ca. 10 Minuten** über die Aussage.
 - Am Ende fasst **jede 8er-Gruppe in 3 Sätzen die Ergebnisse** vor der Klasse zusammen.



¹ <https://www.softguide.de/software-tipps/strukturierte-datenformate>

	JSON	CSV	XML	YAML
Erläuterung (Bedeutung, Herkunft) 				yet another markup language
Aufbau des Formats (Bsp.) 				
Einsatzgebiete 				

1.3 Datenhaltung – aber wie?



Arbeitsauftrag

Ordnen Sie den folgenden Fachbegriff die **richtigen Erklärungen** zu.

9	Inflexibilität	1	Jeder, der autorisiert ist, darf im Mehrbenutzerbetrieb auf die gespeicherten Daten zugreifen. Der Zugriff wird hierbei vom System verwaltet.
5	Daten-unabhängigkeit	2	Die Konsistenz der Daten (d. h. die Übereinstimmung der Datei-Inhalte) kann nur schwer gewährleistet werden. Bei der Änderung einer Größe müssten alle Dateien geändert werden, die diese Größe beinhalten und diese Änderungen müssten so miteinander abgestimmt geschehen, dass nicht verschiedene Programme zum selben Zeitpunkt unterschiedliche Werte derselben Größe sehen können.
1	Mehrfach-zugriff	3	Die Daten müssen in beliebiger Form verknüpfbar sein.
2	Inkonsistenz	4	Die Daten müssen vollständig, korrekt und widerspruchsfrei sein. Daten, die redundant gespeichert sind, müssen das selbe aussagen (Datenkonsistenz).
12	Datenschutz	5	Unabhängigkeit vom Anwendungsprogramm: Die Daten sind anwendungsneutral gespeichert, d. h. unabhängig vom erzeugenden oder benutzenden.
3	Flexibilität	6	Leicht zu erlernende Benutzersprachen ermöglichen sowohl dem professionellen Benutzer (Systementwickler, Programmierer) als auch dem Endbenutzer eine einfache Handhabung der Daten.
11	Daten-Programm-Abhängigkeit	7	Da die Daten jeweils speziell für bestimmte Anwendungen entworfen werden, werden dieselben Daten in verschiedenen Dateien wiederauftauchen (z. B. Namen und Adressen von Lehrern in der Datei für die Lehrer und in der Datei für die Klassen).
6	Benutzer-freundlichkeit	8	Die Zeiten für die Abfrage und für die Verarbeitung müssen kurz sein, ebenso für Änderungen und Ergänzungen des Datenbestandes.
8	Effizienz	9	Da die Daten nicht in ihrer Gesamtheit, sondern nur anwendungsbezogen gesehen werden, ist es in vielen Fällen sehr kompliziert, neue Anwendungen oder Auswertungen vorhandener Daten zu realisieren. Dies gilt insbesondere für Auswertungen, die Daten aus verschiedenen Dateien benötigen.
4	Datenintegrität	10	Jedes Datenelement sollte möglichst nur einmal gespeichert werden, z. B. die Kundenanschrift nur einmal.
7	Redundanz	11	Ändert sich der Aufbau einer Datei oder ihrer Organisationsform, so müssen darauf basierende Programme geändert werden.
10	Redundanz-freiheit	12	Die Daten sind vor unbefugtem Zugriff (Missbrauch) zu schützen.

1.4 Architekturmuster für verteilte Systeme

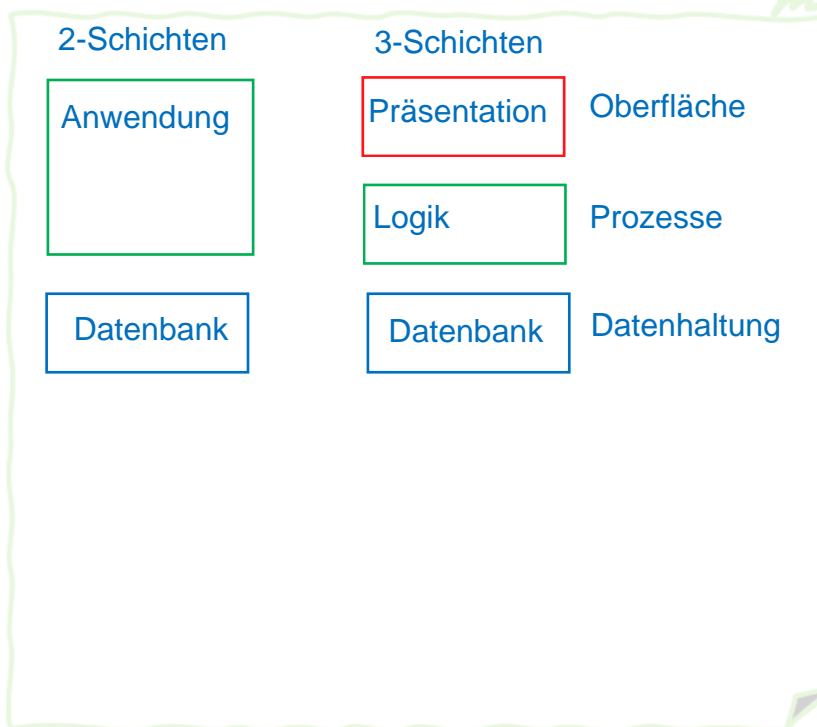


Eine saubere Trennung von Daten und Programm ist bei der Entwicklung verteilter Systeme unumgänglich. Es muss ein Gleichgewicht zwischen Leistung, Verlässlichkeit, Informationssicherheit und Verwaltbarkeit des Systems gefunden werden. Bei Client-Server-Systemen bietet sich eine 3-Schicht-Architektur an.



Arbeitsauftrag

Informieren Sie sich selbstständig mittels Recherche über die **3-Schichtarchitektur** und **skizzieren** Sie diese. **Grenzen** Sie dabei die **2-Schichtarchitektur** von der **3-Schichtarchitektur** ab.



1.5 Phasen der Datenbankentwicklung

Die Entwicklung einer (relationalen) Datenbank verläuft typischerweise in vier aufeinanderfolgenden Phasen. Bringen Sie die vier Phasen in die richtige Reihenfolge (linke Spalte). Erläutern Sie was in dieser Phase geschieht (Mitte) und bestimmen Sie, ob es sich um eine Phase handelt, die bereits von der Wahl einem konkreten Datenbankmanagementsystem abhängig ist (rechte Spalte).



Arbeitsauftrag

Recherchieren Sie nach den 4 Phasen der Datenbankentwicklung und füllen Sie die Tabelle aus.

	Phase	Erläuterung	Abhängigkeit vom DBMS
1	externe phase	Informationsstruktur erfassen	<input type="checkbox"/> Ja <input checked="" type="checkbox"/> Nein
2	Konzeptionelle Phase	Semantisches Modell: <u>Formale</u> u. strukturierte Beschreibung aller relevanten Objekte u. deren Beziehungen untereinander. z.B. ER-Modell (Entity-Relationship-Model)	<input type="checkbox"/> Ja <input checked="" type="checkbox"/> Nein
3	Logische Phase	<u>Relationale Datenbankmodell</u> (rel.Tabellenmodell): aus semantischen Modell wird nach <u>gewissen Regeln</u> ein Relationsmodell entwickelt	<input type="checkbox"/> Ja <input checked="" type="checkbox"/> Nein
4	Physische Phase	Implementierung: Umsetzung in einer Datenbank	<input checked="" type="checkbox"/> Ja <input type="checkbox"/> Nein