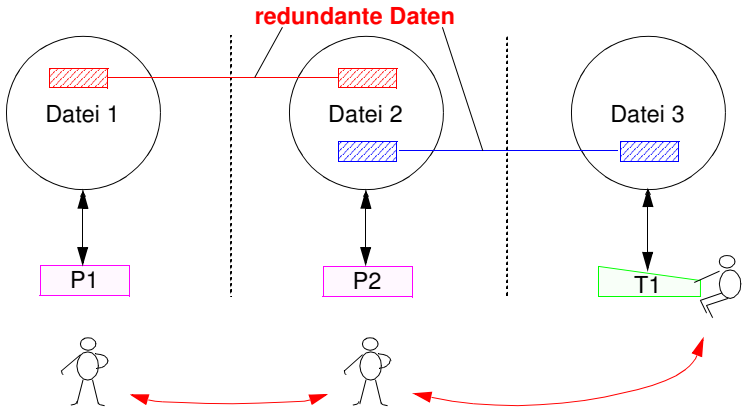
1. Anforderungen und Beschreibungsmodelle

- **Nutzung von Dateisystemen**
- **Grundbegriffe**
- **Anforderungen an ein DBS**
  - Kontrolle über die operationalen Daten
  - Leichte Handhabung der Daten
  - Kontrolle der Datenintegrität
  - Leistung und Skalierbarkeit
  - Hoher Grad an Datenunabhängigkeit
- **Schichtenmodelle für DBS**
  - Statisches Schichtenmodell: „Erklärungsmodell“
  - Schichtenweise Abbildungen
  - Generisches System, Metadaten, Meta-Metadaten
  - Dynamisches Verhalten
- **Wo steckt die Bedeutung?**
  - Die ganze Wahrheit
  - Meta, Meta-Meta, Meta-Meta-Meta
- **Drei-Schema-Architektur nach ANSI-SPARC**
  - Externes, konzeptionelles und internes Schema
  - Beschreibungsebenen für DB-Anwendungen
- **Dynamischer Ablauf von DBS-Operationen**

Nutzung von Dateisystemen

- **Permanente Datenhaltung innerhalb von BS-Dateien:**
  Warum keine direkte Nutzung von Dateien zur Datenhaltung in IS?
- **Betriebssystem/Dateisystem bietet Funktionen für**
  - Erzeugen / Löschen von Dateien
  - Zugriffsmöglichkeiten auf Blöcke/Sätze der Datei
  - einfache Operationen zum Lesen/Ändern/Einfügen/Löschen von Sätzen (dynamisches Wachstum)



Kommunikation notwendig für Änderungen

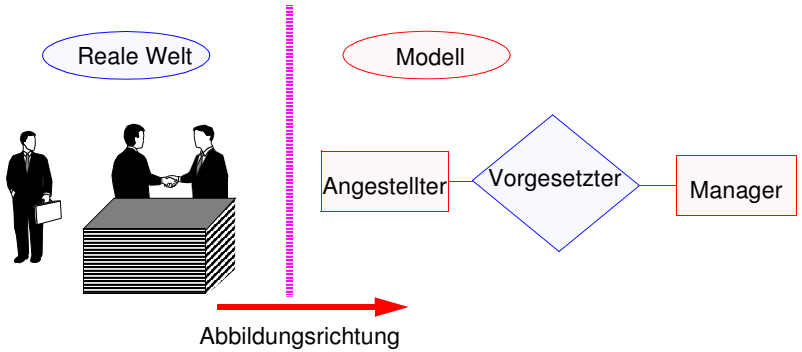
- **Probleme/Nachteile**
  - Datenredundanz und Inkonsistenz
  - Inflexibilität
  - **Mehrbenutzerbetrieb, Fehlerfall**
  - **Integritätssicherung**
  - Missbrauch der Daten

➥ **Wer ist verantwortlich?**

- **DBS nutzt gewöhnlich Dateien** des Betriebssystems / Dateisystems zur permanenten Datenhaltung (**Externspeicherabstraktion**)

Grundbegriffe

- **Datenbank als Abbildung einer Miniwelt**
  - Vorgänge und Sachverhalte werden als **gedankliche Abstraktionen (Modelle) der Miniwelt** erfasst und als Daten (Repräsentationen von Modellen) in der Datenbank gespeichert
  - Daten beziehen sich nur auf solche Aspekte der Miniwelt, die **für die Zwecke der Anwendung relevant** sind
  - Eine DB ist **integritätserhaltend** (bedeutungstreu), wenn **ihre Objekte Modelle einer gegebenen Miniwelt** repräsentieren



- **Datenmodell und DB-Schema**
  - Datenmodell (Typen, Operatoren, Konsistenzbedingungen) legt Regeln fest, nach denen die Objekte von DBs (für die Repräsentation beliebiger Miniwelten) erzeugt und verändert werden
    (Konstruktionsregeln für die Zustandsräume der Modelle)
  - DB-Schema legt die Ausprägungen der Objekte fest, welche die DB für eine bestimmte Miniwelt einnehmen kann
    (Zustandsraum der Modelle einer Miniwelt)

Grundbegriffe (2)

- **Beschreibung und Handhabung der Daten**
  - Daten müssen interpretierbar sein
  - Sie müssen bei allen am Austausch beteiligten Partnern (Systemen, Komponenten) die Ableitung derselben Information erlauben

    ➥ Rolle des DB-Schemas

Schema

ANGESTELLTER

Satztyp (Relation)

Ausprägungen

PNR	NAME	TAETIGKEIT	GEHALT	ALTER
496	PEINL	PFOERTNER	2100	63
497	KINZINGER	KOPIST	2800	25
498	MEYWEG	KALLIGRAPH	4500	56

  - Interpretierbarkeit der Daten muss zeitinvariant sein
  - Einsatzspektrum verlangt **generische Vorgehensweise**
    - Beschreibung der zulässigen DB-Zustände
    - Beschreibung der zulässigen Zustandsübergänge (generische Operatoren)

- **Anwendungsprogrammier-Schnittstelle (API)**
  - Operatoren zur Definition von Objekttypen (Beschreibung der Objekte)

    ➥ DB-Schema: Welche Objekte sollen in der DB gespeichert werden?
  - Operatoren zum Aufsuchen und Verändern von Daten

    ➥ AW-Schnittstelle: Wie erzeugt, aktualisiert und findet man DB-Objekte?
  - Definition von Integritätsbedingungen (*Constraints*)

    ➥ Sicherung der Qualität: Was ist ein akzeptabler DB-Zustand?
  - Definition von Zugriffskontrollbedingungen

    ➥ Maßnahmen zum Datenschutz: Wer darf was?

Anforderungen an ein DBS

1. Kontrolle über die operationalen Daten

- **Alle Daten können/müssen gemeinsam benutzt werden**
  - keine verstreuten privaten Dateien
  - Querauswertungen aufgrund inhaltlicher Zusammenhänge
  - symmetrische Organisationsformen
    (keine Bevorzugung einer Verarbeitungs- und Auswertungsrichtung)
  - Entwicklung neuer Anwendungen auf der existierenden DB
  - Erweiterung/Anpassung der DB (Änderung des Informationsbedarfs)

  ➥ ***Anwendungsneutralität beim DB-Entwurf*:**
  Was zeichnet ein gutes DB-Schema aus?

- **Eliminierung der Redundanz**
  - keine wiederholte Speicherung in unterschiedlicher Form
    für verschiedene Anwendungen
  - Vermeidung von Inkonsistenzen
  - zeitgerechter Änderungsdienst,
    keine unterschiedlichen Änderungsstände

  ➥ ***Redundanzfreiheit aus der Sicht der Anwendungen***

- **Datenbankadministrator (DBA):**
  zentrale Verantwortung für die operationalen Daten[1]

1. Die weitaus meisten Daten werden auch physisch zentralisiert verwaltet. Etwa 2/3 aller weltweit relevanten wirtschaftlichen Daten werden im EBCDIC-Format auf Rechnern der S/390-, BS2000- und AS/400-Architektur gespeichert. 60% aller vom Web aufrufbarer Daten befinden sich auf Mainframes. Es dominieren Datenbanken wie DB2, IMS und VSAM (W. G. Spruth).

2. Leichte Handhabbarkeit der Daten

- **Einfache Datenmodelle**
  - Beschreibung der logischen Aspekte der Daten
  - Benutzung der Daten ohne Bezug auf systemtechnische Realisierung

- **Logische Sicht der Anwendung**
  - zugeschnitten auf ihren Bedarf
  - lokale Sicht auf die DB

- **Leicht erlernbare Sprachen**
  - deskriptive Problemformulierung
  - hohe Auswahlmächtigkeit
  - Unterstützung der Problemlösung des Anwenders im Dialog

- **Durchsetzung von Standards**
  - unterschiedliche DBS bieten einheitliche Schnittstelle
  - Portierbarkeit von Anwendungen
  - erleichterter Datenaustausch

- **Erweiterung der Benutzerklassen**
  - Systempersonal
  - Anwendungsprogrammierer
  - Anspruchsvolle Laien
  - Parametrische Benutzer/Gelegentliche Benutzer

Relationenmodell – Beispiel

DB-Schema

FB

FBNR	FBNAME	DEKAN

PROF

PNR	PNAME	FBNR	FACHGEB

STUDENT

MATNR	SNAME	FBNR	STUDBEG

PRÜFUNG

PNR	MATNR	FACH	DATUM	NOTE

Ausprägungen

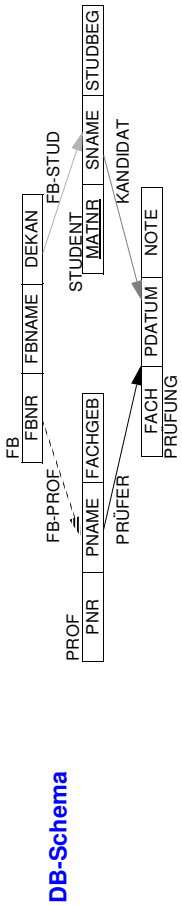
FB	FBNR	FBNAME	DEKAN
	FB 9	WIRTSCHAFTSWISS	4711
	FB 5	INFORMATIK	2223

PROF	PNR	PNAME	FBNR	FACHGEB
	1234	HÄRDER	FB 5	DATENBANKSYSTEME
	5678	WEDEKIND	FB 9	INFORMATIONSSYSTEME
	4711	MÜLLER	FB 9	OPERATIONS RESEARCH
	6780	NEHMER	FB 5	BETRIEBSSYSTEME

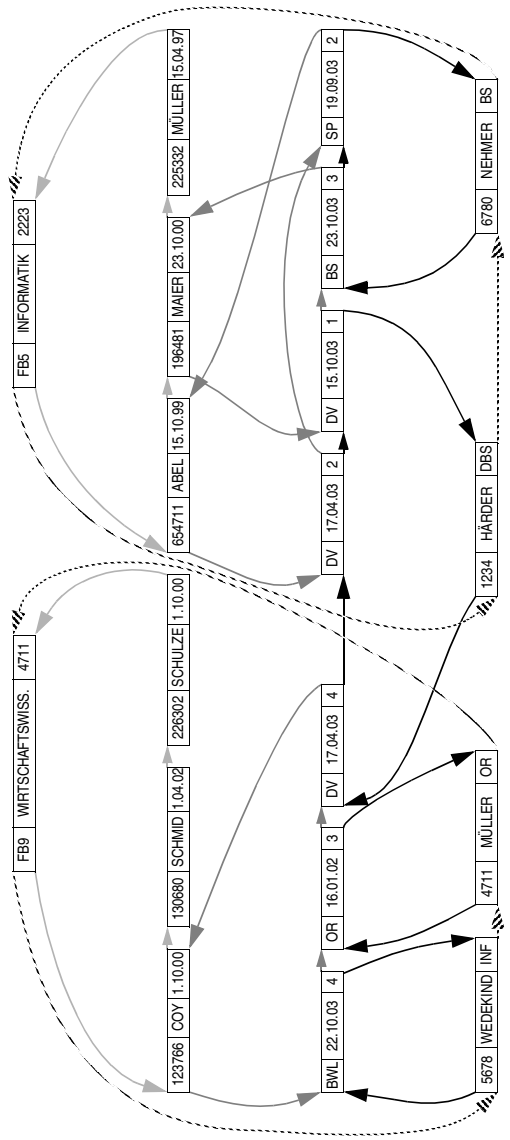
STUDENT	MATNR	SNAME	FBNR	STUDBEG
	123 766	COY	FB 9	1.10.00
	225 332	MÜLLER	FB 5	15.04.97
	654 711	ABEL	FB 5	15.10.99
	226 302	SCHULZE	FB 9	1.10.00
	196 481	MAIER	FB 5	23.10.00
	130 680	SCHMID	FB 9	1.04.02

PRÜFUNG	PNR	MATNR	FACH	PDATUM	NOTE
	5678	123 766	BWL	22.10.03	4
	4711	123 766	OR	16.01.02	3
	1234	654 711	DV	17.04.03	2
	1234	123 766	DV	17.04.03	4
	6780	654 711	SP	19.09.03	2
	1234	196 481	DV	15.10.03	1
	6780	196 481	BS	23.10.03	3

Netzwerkmodell – Beispiel



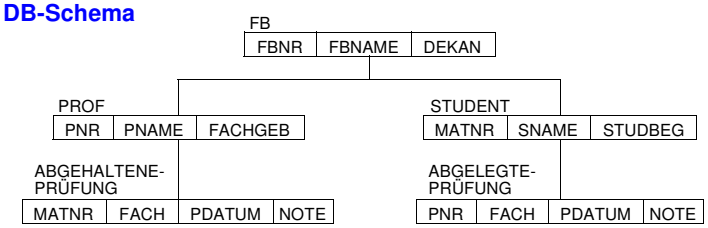
DB-Schema



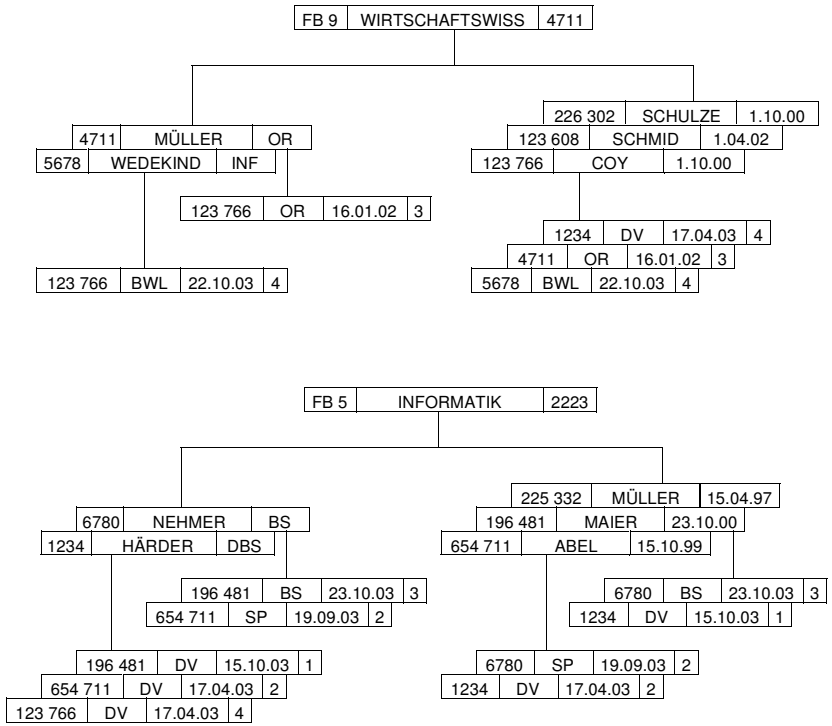
Ausprägungen

Hierarchisches Datenmodell – Beispiel

DB-Schema

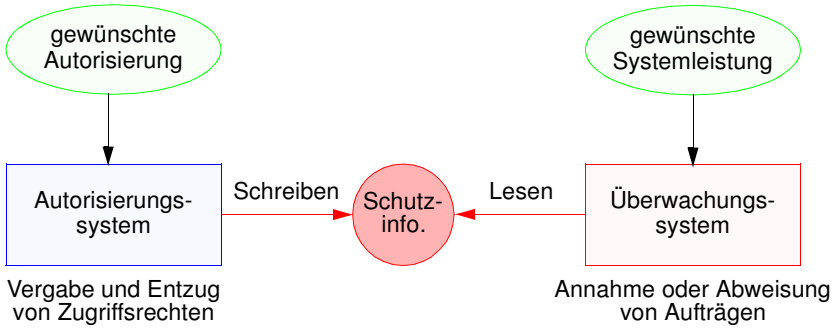


Ausprägungen

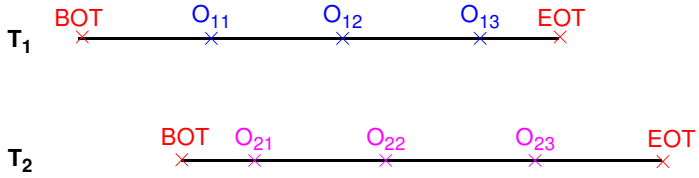


3. Kontrolle der Datenintegrität

- **Automatisierte Zugriffskontrollen (Datenschutz)**
  - separat für jedes Datenobjekt
  - unterschiedliche Rechte für verschiedene Arten des Zugriffs
  - **Idealziel**: „least privilege principle“



- **Erhaltung der logischen Datenintegrität (system enforced integrity)**
  - Beschreibung der „Richtigkeit“ von Daten durch Prädikate und Regeln
  - „Qualitätskontrollen“ bei Änderungsoperationen
  - aktive Maßnahmen des DBS erwünscht (ECA-Regeln)



BOT: Begin of Transaction EOT (Commit): End of Transaction

O_{ij} : DB-Operation; Lese- und Schreiboperationen auf DB-Daten

3. Kontrolle der Datenintegrität (Forts.)

- **Transaktionskonzept** (Durchsetzung der ACID-Eigenschaften[2])
  - Schema-Konsistenz (**C**) aller DB-Daten wird bei Commit erzwungen
  - ACID impliziert Robustheit, d. h., DB enthält nur solche Zustände, die explizit durch erfolgreich abgeschlossene TA erzeugt wurden
    - **D**auerhaftigkeit (Persistenz): Effekte von abgeschlossenen TA gehen nicht verloren
    - **A**tomarität (Resistenz): Zustandsänderungen werden entweder, wie in der TA spezifiziert, vollständig durchgeführt oder überhaupt nicht
  - Im Mehrbenutzerbetrieb entsteht durch nebenläufige TA ein Konkurrenzverhalten (concurrency) um gemeinsame Daten, d. h., TA geraten in Konflikt

    ➥ **I**solationseigenschaft: TA-Konflikte sind zu verhindern oder aufzulösen

- **Erhaltung der physischen Datenintegrität**
  - Periodisches Erstellen von Datenkopien
  - Führen von Änderungsprotokollen für den Fehlerfall (Logging)
  - Bereitstellen von Wiederherstellungsalgorithmen im Fehlerfall (Recovery)

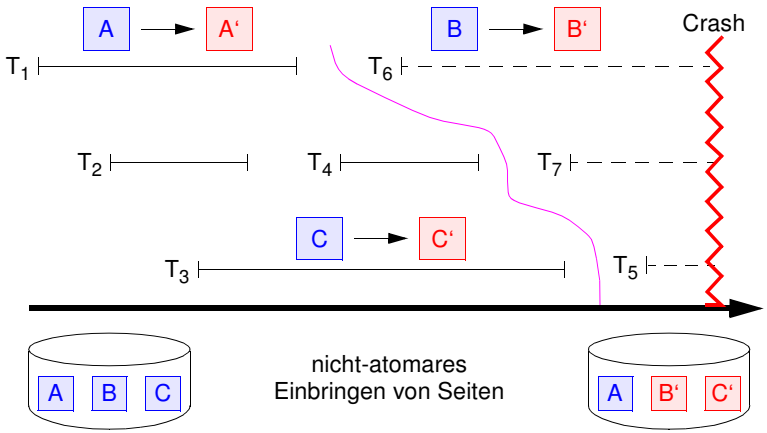
    ➥ **Garantie** nach erfolgreichem Neustart:
    jüngster transaktionskonsistenter DB-Zustand

- **Notwendigkeit des kontrollierten Mehrbenutzerbetriebs**
  - logischer Einbenutzerbetrieb für jeden von n parallelen Benutzern (Leser + Schreiber)
  - geeignete Synchronisationsmaßnahmen zur gegenseitigen Isolation
  - angepasste Synchronisationseinheiten (z. B. Sperrgranulate) mit abgestuften Zugriffsrechten

    ➥ **Ziel**: möglichst geringe gegenseitige Behinderung

2. „May all your transactions commit and never leave you in doubt“ (J. Gray)

Physische Datenintegrität – jüngster transaktionskonsistenter Zustand)

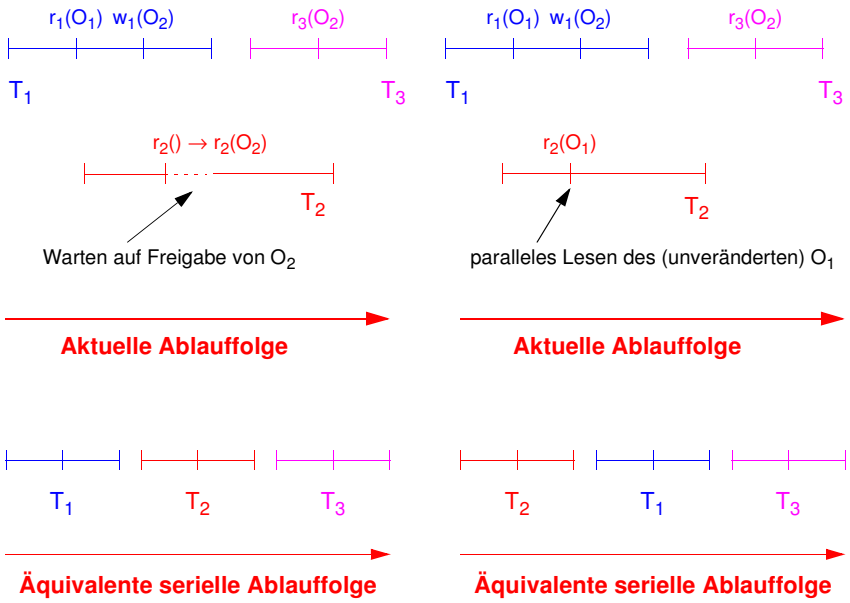


- **DBMS garantiert physische Datenintegrität**
  - Bei jedem Fehler (z. B. Ausfall des Rechners, Crash des Betriebssystems oder des DBMS, Fehlerhaftigkeit einzelner Transaktionsprogramme) wird eine „korrekte“ Datenbank rekonstruiert
  - Nach einem (Teil-)Crash ist immer der jüngste transaktionskonsistente Zustand der DB zu rekonstruieren, in dem alle Änderungen von Transaktionen enthalten sind, die vor dem Zeitpunkt des Fehlers erfolgreich beendet waren (T_1 bis T_4) und sonst keine
  - automatische Wiederherstellung bei Restart (Wiederanlauf) des Systems

- **Maßnahmen beim Wiederanlauf (siehe auch Beispiel)**
  - Ermittlung der beim Crash aktiven Transaktionen (T_5, T_6, T_7)
  - Wiederholen (REDO) der Änderungen von abgeschlossenen Transaktionen, die vor dem Crash nicht in die Datenbank zurückgeschrieben waren (A $\rightarrow$ A‘)
  - Rücksetzen (UNDO) der Änderungen der aktiven Transaktionen in der Datenbank (B‘ $\rightarrow$ B)

Logischer Einbenutzerbetrieb

- Beim **logischen Einbenutzerbetrieb** hat jede der parallel aktiven Transaktionen den Eindruck, als liefe sie alleine ab, d. h., logisch bilden alle Transaktionen eine serielle Ablauffolge
- **Synchronisationskomponente** des DBMS umfasst alle Maßnahmen zur Sicherstellung der Ablaufintegrität (Isolation der parallelen Transaktionen)
- **Formale Definition:** Eine parallele Ablauffolge von Transaktionen ist genau dann korrekt synchronisiert, wenn es eine zu dieser Ablauffolge äquivalente (bezüglich ihrer Lese- und Schreibabhängigkeiten (r, w)) serielle Ablauffolge gibt, so dass jede Transaktion T_i in der seriellen Reihenfolge dieselben Werte liest und schreibt wie im parallelen Ablauf. (Dabei ist jede Permutation der T_i-Folge gleichermaßen zulässig, siehe Beispiel).



4. Leistung und Skalierbarkeit

- **DBS-Implementierung gewährleistet**
  - **Effizienz** der Operatoren (möglichst geringer Ressourcenverbrauch)
  - **Verfügbarkeit** der Daten (Redundanz, Verteilung usw.)
- **Ausgleich von Leistungsanforderungen, die im Konflikt stehen**
  - globale Optimierung durch den DBA (Rolle des internen Schemas)
  - ggf. Nachteile für einzelne Anwendungen
- **Effizienz des Datenzugriffs**
  - Zugriffsoptimierung durch das DBS, nicht durch den Anwender
  - Auswahl von Zugriffspfaden durch den DBA
    ➥ idealerweise durch das DBS
- **Leistungsbestimmung**
  - Maßzahlen für Leistung
    - Durchsatz: Anzahl abgeschlossener TA pro Zeiteinheit (meist Sekunde)
    - Antwortzeit: Zeitbedarf für die Abwicklung einer TA
  - Rolle von Benchmarks[3]: TPC-C, TPC-H, TPC-W, TPC-R, . . .
- **Skalierbarkeit**
  - Software- und Hardware-Architektur[4] sollen hinsichtlich des DBS-Leistungsverhaltens automatisch durch Hinzufügen von Ressourcen (CPU, Speicher) skalieren
    - **Scaleup:** bei Wachstum der Anforderungen (DB-Größe, TA-Last)
    - **Speedup:** zur Verringerung der Antwortzeit

3. **T**ransaction **P**rocessing **C**ouncil: www.tpc.org
4. Der S/390-Sysplex stellt eine Clustering-Technologie dar, bei der es möglich ist, Standardanwendungen wie DB2, CICS, IMS und Unix-System-Services von 2 CPUs auf 100 CPUs zu skalieren, mit einem Leistungsabfall im einstelligen Prozentbereich (im Vergleich zum linearen Wachstum).

5. Hoher Grad an Datenunabhängigkeit

- **Konventionelle Anwendungsprogramme (AP) mit Dateizugriff**
  - Nutzung von Kenntnissen der Datenorganisation und Zugriffstechnik
  - gutes Leistungsverhalten, aber . . . ?
- **Datenabhängige Anwendungen sind äußerst unerwünscht**
  - Rolle des Datenmodells:
    Vergleiche relationales und hierarchisches Datenmodell
  - Verschiedene Anwendungen brauchen verschiedene Sichten auf dieselben Daten
  - Änderungen im Informationsbedarf sowie bei Leistungsanforderungen (sehr häufig) erzwingen Anpassungen bei Speicherungsstrukturen und Zugriffsstrategien

    ➥ deshalb: *möglichst starke Isolation der APs von den Daten*
    sonst: extremer Wartungsaufwand für die APs
- **Realisierung verschiedener Arten von *Datenunabhängigkeit*:**
  - **Geräte**unabhängigkeit
  - **Speicherungsstruktur**unabhängigkeit

    ➥ **Minimalziel: physische Datenunabhängigkeit**
    (durch das Betriebssystem/Datenbanksystem)
  - **Zugriffspfad**unabhängigkeit
  - **Datenstruktur**unabhängigkeit

    ➥ **logische Datenunabhängigkeit**
    (vor allem durch das Datenmodell!)

Der „Religionskrieg“ – Welche Abstraktionsebene gewinnt?

- **Netzwerkmodell**
  - Je komplexer die Datenstruktur, desto besser
  - Aber: sehr einfache Operationen

  ➥ **Verzeigerung, Navigation, Satzorientierung**
- **Relationenmodell**
  - „data structure of spartan simplicity“ (E. F. Codd)
  - Operationen mit Hülleneigenschaft
  - Jeder Schnörkel benötigt zusätzliche Operationen
  - Deshalb: Einfachheit ist das Geheimnis für Datenunabhängigkeit

  ➥ **Wertbasierung, Deklarativität, Mengenorientierung**

Strukturierte Programmierung und Datenunabhängigkeit

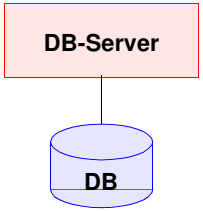
- **E. W. Dijkstra’s Ideen**
  - „Notes on Sturctured Programming“ (EWD249, 1969)
  - „The goto statement considered harmful“ (Comm. ACM, 1968)

  ➥ Kampfesruf übersetzt in die Welt der Datenbanken:

  Zeiger (Pointer) sind von Übel!
- **ab 1971: D. L. Parnas publiziert seine Ideen zu**
  - „Information Hiding“
  - hierarchische Strukturierung von SW-Systemen
  - ...

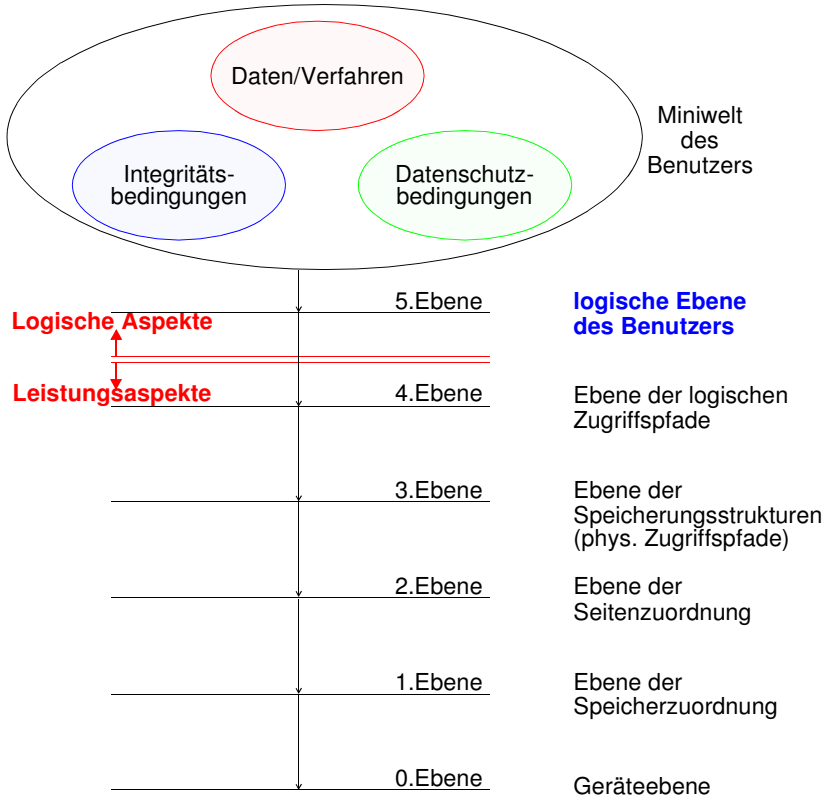
Auf dem Weg zum Schichtenmodell

- **Die durch das Relationenmodell (Sprachebene) postulierte Unabhängigkeit musste auch systemtechnisch umgesetzt werden**
- **Monolithischer Ansatz?**
  - Q1: Select B.Titel, P.E-Jahr, A.Name
    From Bücher B, Autoren A
    Where B.Autor = A.Autor And A.Name = „S*“ And B.Fach = „DBS“



  - Schnittstelle zum Externspeicher: Lesen und Schreiben von Seiten (DB ist ein sehr langer Bitstring!)
- **Permanente Evolution erwartet (~1980)**
  - Lange Lebenszeit eines DBMS > 30 Jahre
  - wachsender Informationsbedarf: Objekttypen, Integritätsbedingungen, ...
  - neue Speicherungsstrukturen und Zugriffsverfahren, ...
  - schnelle Änderungen der Technologien, Speicher, ...
- **Aber keine Revolution!**
  - Raum und Zeit, Datenströme?
  - unstrukturierte und semistrukturierte Dokumente? ...
- **Wichtige Herausforderungen**
  - Logische und Physische Datenunabhängigkeit
    - Möglichst starke Trennung von AW-Programmen und Daten
    - Isolation der Schnittstelle von allen Änderungen im DBS
  - Kapselung und hierarchische Strukturierung im System

Ebenen beim Entwurf eines DBS[5]



- **„outside-in“-Ansatz** (*top-down*)
  - Die Ausdrucksmächtigkeit des Datenmodells und seine Konzepte sowie die postulierten Betriebseigenschaften bestimmen die Anforderungen, die an das zu entwerfende DBS zu stellen sind
  - Beim Entwurf erfolgt eine mehrstufige Strukturverfeinerung, bis die konkrete Implementierungsstruktur abgeleitet ist

5. „Eine Hauptaufgabe der Informatik ist systematische Abstraktion“ (H. Wedekind)

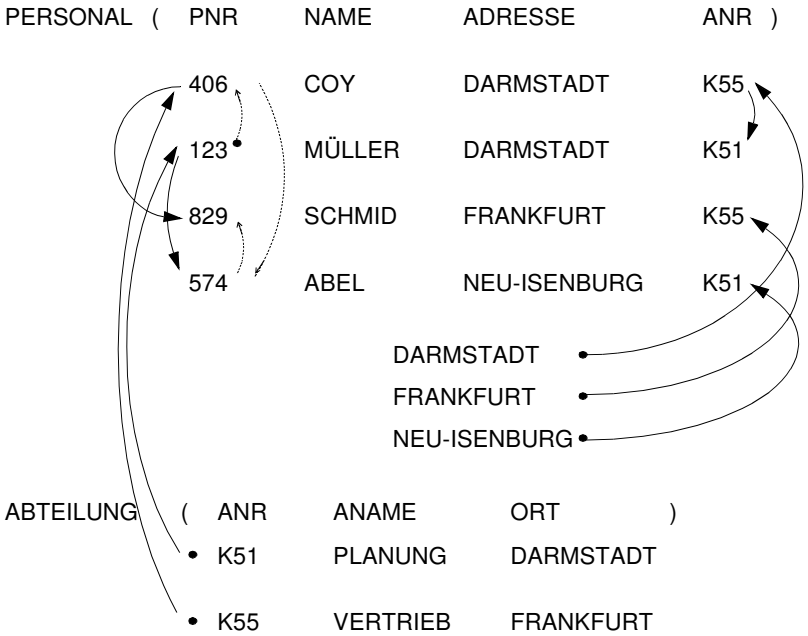
Verschiedene Sichten auf DBS-Daten

- **Logischen Datenstrukturen eines Anwendungsbeispiels**

PERSONAL (	PNR	NAME	ADRESSE	ANR)
	406	COY	DARMSTADT	K55
	123	MÜLLER	DARMSTADT	K51
	829	SCHMID	FRANKFURT	K55
	574	ABEL	NEU-ISENBURG	K51

ABTEILUNG (	ANR	ANAME	ORT)
	K51	PLANUNG	DARMSTADT
	K55	VERTRIEB	FRANKFURT

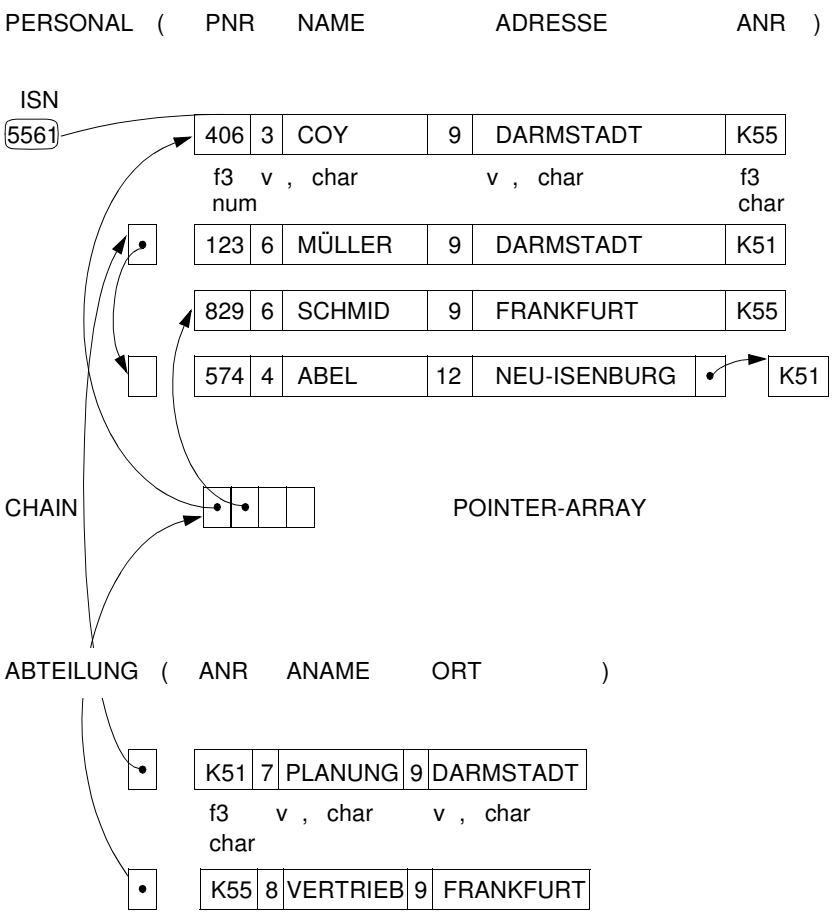
- **Sicht auf die logischen Zugriffspfade**



Logische Zugriffspfade: 1. OWNER – MEMBER
2. Sortierreihenfolge PNR ASC
3. Search Key (Invertierung ADRESSE)

Verschiedene Sichten auf DBS-Daten (2)

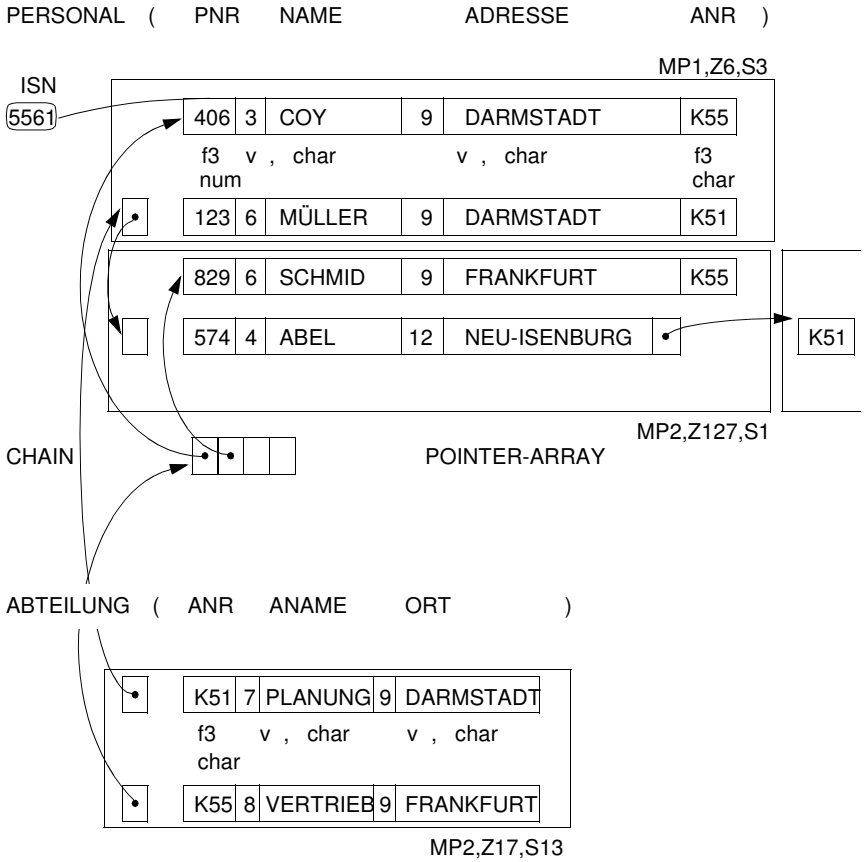
- **Sicht auf die Speicherungsstrukturen**



Speicherungsstrukturen: 1. Formate
2. Datentypen
3. Implementierungstechniken

Verschiedene Sichten auf DBS-Daten (3)

- **Sicht auf die Speicherzuordnungsstrukturen**



Speicherzuordnungsstrukturen: 1. physische Blocklänge
2. spanned record facility

Gerätemerkmale: 1. Eigenschaften der Speichermedien
2. Magnetplatten-Zuordnungen

Schichtenmodelle für DBS

- **Ziel: Architektur eines datenunabhängigen DBS**

- **Systementwurf**
  - Was sind die geeigneten Beschreibungs- und Kommunikationstechniken? Sie sind notwendigerweise informal.
  - Was ist auf welcher Beschreibungsebene sichtbar?
    Es ist angemessene Abstraktion erforderlich![6]
  - Wie kann eine Evolution des Systems erfolgen?
    Es muss eine Kontrolle der Abhängigkeiten erfolgen!

- **Aufbau in Schichten:**
  - „günstige Zerlegung“ des DBS in „nicht beliebig viele“ Schichten
  - optimale Bedienung der Aufgaben der darüberliegenden Schicht
  - implementierungsunabhängige Beschreibung der Schnittstellen

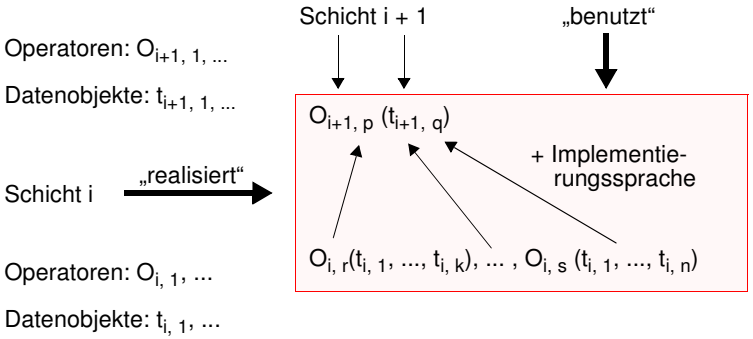
  ➥ Es gibt keine Architekturlehre für den Aufbau großer SW-Systeme

- **Empfohlene Konzepte:**
  - Geheimnisprinzip (*Information Hiding*)
  - Trennung der Belange (*Separation of Concerns*)
  - hierarchische Strukturierung
  - generische Auslegung der Schnittstellen:
    Nur bestimmte Objekttypen mit charakteristischen Operationen sind vorgegeben, jedoch nicht ihre anwendungsbezogene Spezifikation und Semantik

6. „Die durch Abstraktion entstandenen Konstrukte der Informatik als Bedingungen möglicher Information sind zugleich die Bedingungen der möglichen Gegenstände der Information in den Anwendungen“ (H. Wedekind in Anlehnung an eine Aussage Kants aus der „Kritik der reinen Vernunft“) Vereinfacht ausgedrückt: Informatiker erfinden (konstruieren) abstrakte Konzepte; diese ermöglichen (oder begrenzen) wiederum die spezifischen Anwendungen.

Schichtenmodelle für DBS (2)[7]

- **Aufbauprinzip:**



- **„benutzt“-Relation:**

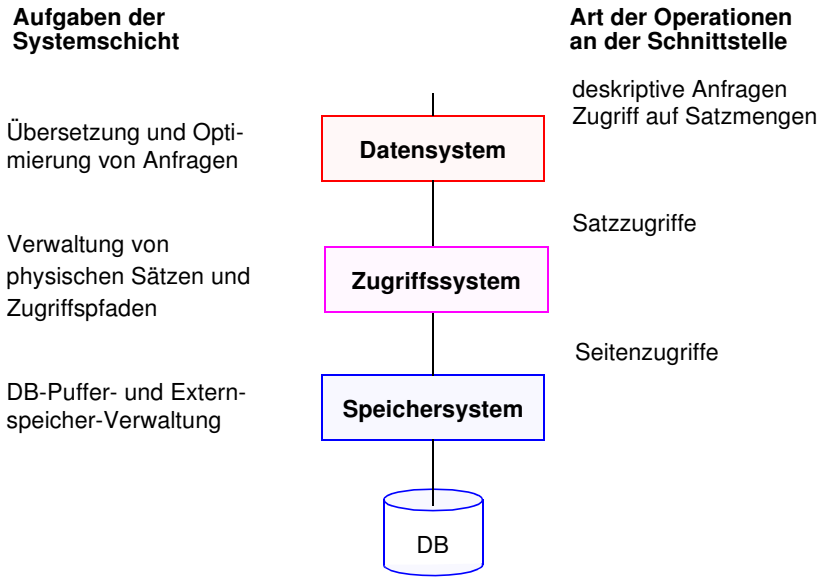
  A **benutzt** B, wenn A B aufruft und die korrekte Ausführung von B für die vollständige Ausführung von A notwendig ist

- **Anzahl der Schichten**
  - n = ?
  - Entwurfskomplexität/Schicht fällt mit wachsendem n
  - Laufzeitaufwand des DBS steigt mit wachsendem n

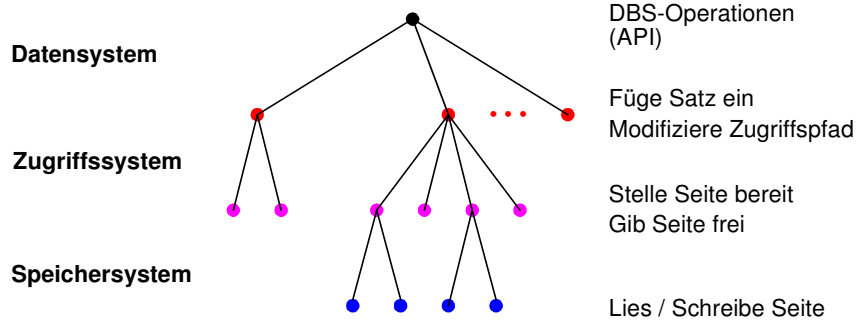
7. Härder, T., Rahm, E.: Datenbanksysteme – Konzepte und Techniken der Implementierung, Springer-Verlag, 2001, Kap. 1

Schichtenmodelle für DBS (3)

- **Vereinfachtes Schichtenmodell**



- **Dynamischer Kontrollfluss einer Operation an das DBS**



Schichtenmodelle für DBS (4)

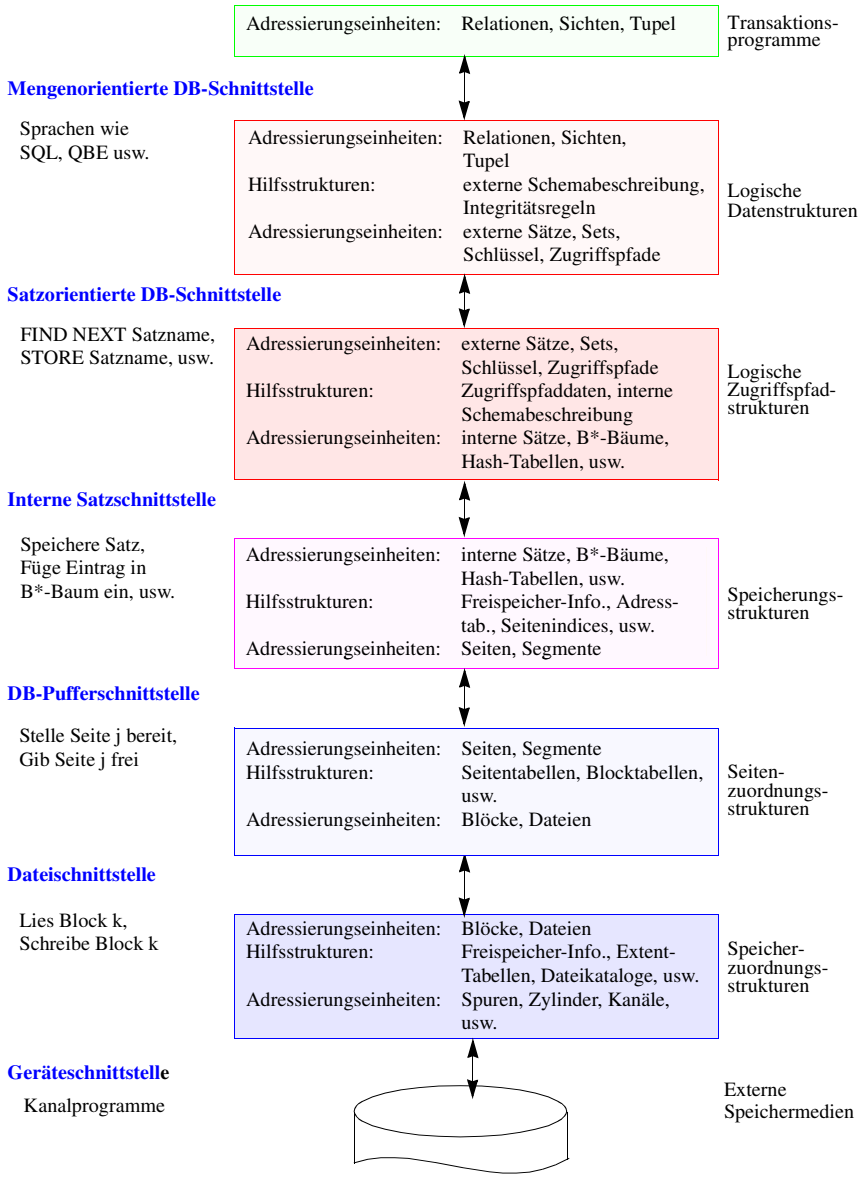
- Vorteile als Konsequenzen der Nutzung **hierarchischer Strukturen und der „benutzt“-Relation**
  - Höhere Ebenen (Systemkomponenten) werden einfacher, weil sie tiefere Ebenen (Systemkomponenten) benutzen können
  - Änderungen auf höheren Ebenen sind ohne Einfluss auf tieferen Ebenen
  - Höhere Ebenen können abgetrennt werden, tiefere Ebenen bleiben trotzdem funktionsfähig
  - Tiefere Ebenen können getestet werden, bevor die höheren Ebenen lauffähig sind

- **Jede Hierarchieebene kann als abstrakte oder virtuelle Maschine aufgefasst werden**
  - Programme der Schicht i benutzen als abstrakte Maschine die Programme der Schicht i-1, die als Basismaschine dienen
  - Abstrakte Maschine der Schicht i dient wiederum als Basismaschine für die Implementierung der abstrakten Maschine der Schicht i+1

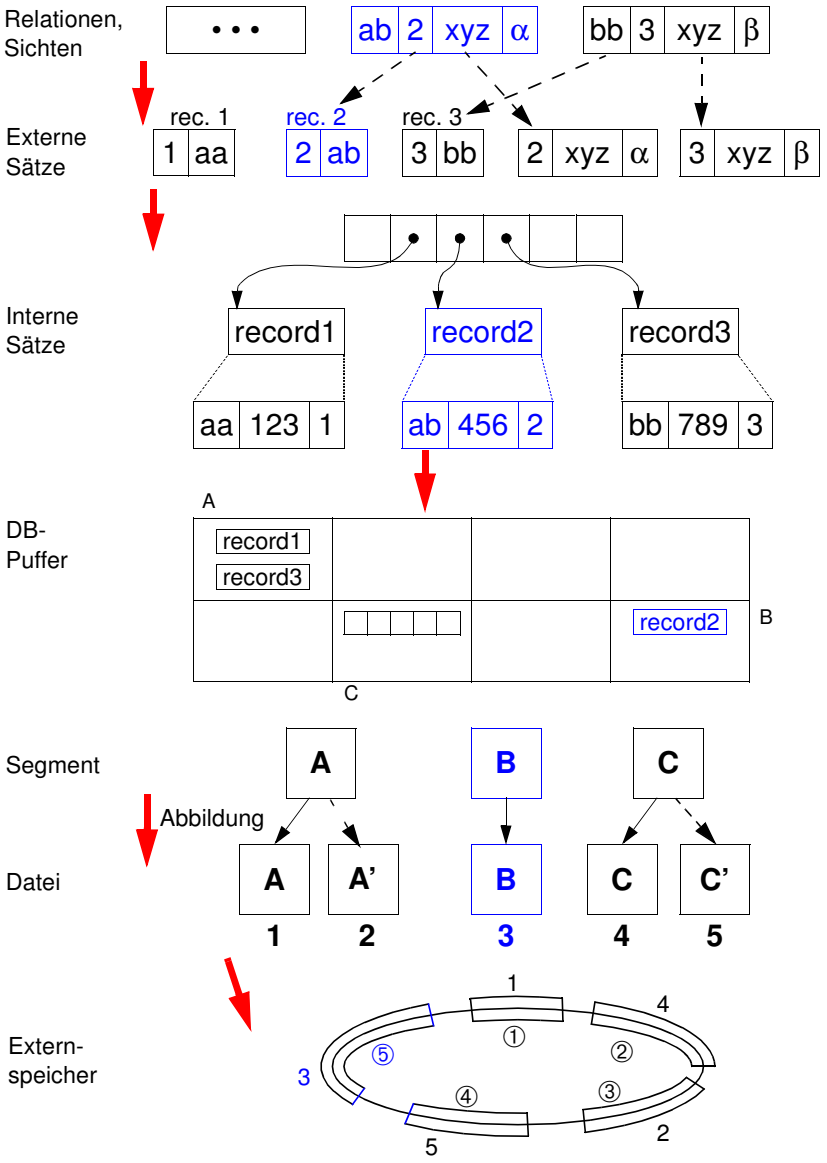
- **Eine abstrakte Maschine entsteht aus der Basismaschine durch Abstraktion**
  - Einige Eigenschaften der Basismaschine werden verborgen
  - Zusätzliche Fähigkeiten werden durch Implementierung höherer Operationen für die abstrakte Maschine bereitgestellt

- Programme einer bestimmten Schicht können die der nächsten tieferen Schicht genau so benutzen**, als sei die untere Schicht Hardware**

Statisches Modell eines Datenbanksystems



Schichtenweise Abbildungen in einem DBS

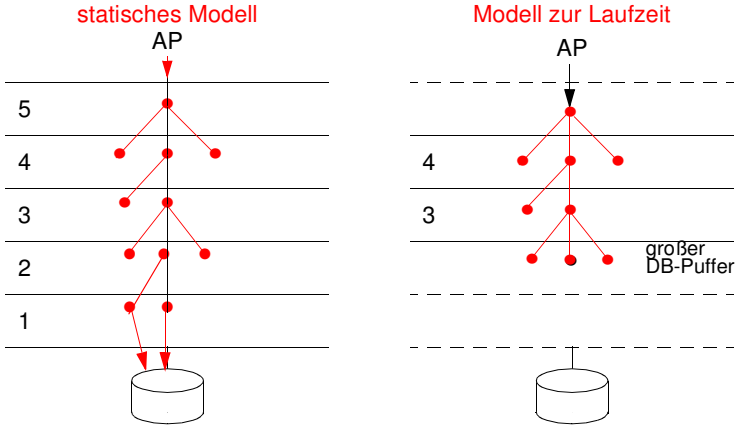


n-Schichtenmodell – Rolle von n

- **n = 1: Monolith**
- **Wachsendes n (n < 10)**
  - Reduktion der Komplexität der einzelnen Schichten (leichtere Systemevolution)
  - Leistungsverluste, da mehr Schnittstellen zu überqueren sind.[8] Bei jedem Übergang:
    - Kapselung/Parameterüberprüfung
    - Datentransport
      - Kopiervorgänge nach oben!
      - Propagieren von Änderungen nach unten!
    - nicht-lokale Fehlerbehandlung ist schwieriger (Verstehen von Fehlermeldungen?)
  - Abnehmende Möglichkeit der Optimierung

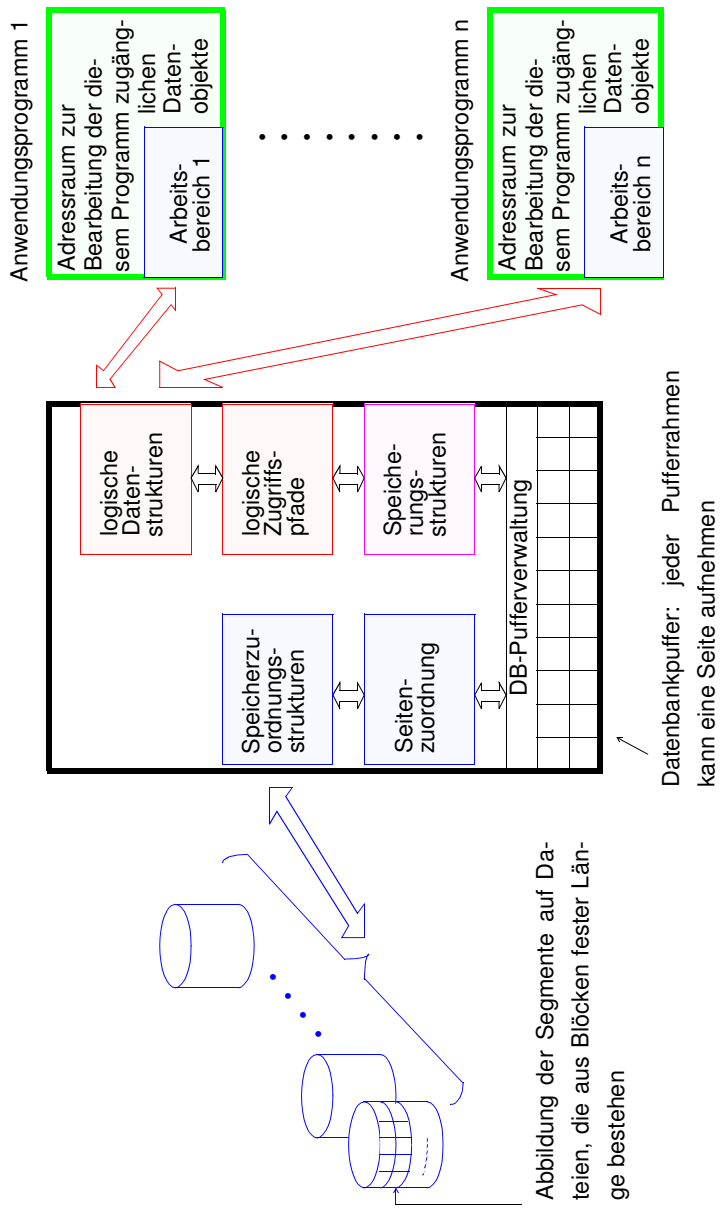
➥ Kompromiss bei der Wahl von n!

- **n = 5**



8. Für DBS gilt besonders: „Leistung ist nicht alles, aber ohne Leistung ist alles nichts!“

Schichtenmodell – Laufzeitaspekte



Architektur eines DBS – weitere Ziele

- **Datenunabhängigkeit im Überblick**

Ebene	Was wird verborgen?
Logische Datenstrukturen	Positionsanzeiger und explizite Beziehungskonstrukte im Schema
Logische Zugriffspfade	Zahl und Art der physischen Zugriffspfade; interne Satzdarstellung
Speicherungsstrukturen	Pufferverwaltung; Recovery-Vorkehrungen
Seitenzuordnungsstrukturen	Dateiabbildung, Recovery-Unterstützung durch das BS
Speicherzuordnungsstrukturen	Technische Eigenschaften und Betriebsdetails der externen Speichermedien

- **Entwurfsziel:**
  DBS sollen von ihrem Aufbau und ihrer Einsatzorientierung her in hohem Maße **generische Systeme** sein. Sie sind so zu entwerfen, dass sie flexibel durch Parameterwahl und ggf. durch Einbindung spezieller Komponenten für eine vorgegebene Anwendungsumgebung zu konfigurieren sind

- **Visionen und Forschungsziele**
  (verlangen perfekte System- und Datenbeschreibungen)
  - **Autonomic Computing:**
    Systeme, insbesondere DBS, werden so komplex, dass sie sich selbst administrieren und optimieren sollten!?
  - **Organic Computing:**
    Hier geht es vor allem um selbstorganisierende Systeme (z. B. im Internet), die sich den jeweiligen Anforderungen ihrer Umgebung dynamisch anpassen sollen. Buzzwords für ihre Eigenschaften sind selbstkonfigurierend, selbstoptimierend, selbstheilend und selbstschützend

Architektur eines DBS – weitere Komponenten

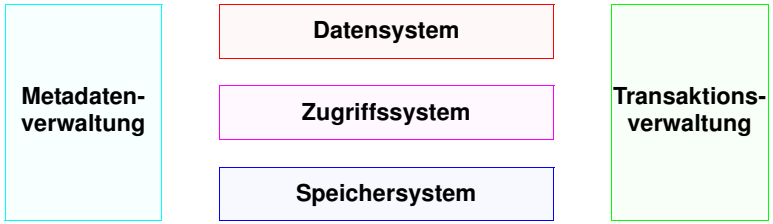
- **Rolle der Metadaten**
  - Metadaten enthalten Informationen über die zu verwaltenden Daten
  - Sie beschreiben also diese Daten (Benutzerdaten) näher hinsichtlich Inhalt, Bedeutung, Nutzung, Integritätsbedingungen, Zugriffskontrolle usw.
  - Die Metadaten lassen sich unabhängig vom DBVS beschreiben (siehe internes, konzeptionelles und externes Schema)

  ➥ Dadurch erfolgt das „Zuschneidern“ eines DBS auf eine konkrete Einsatzumgebung. Die Spezifikation, Verwaltung und Nutzung von Metadaten bildet die Grundlage dafür, dass DBS hochgradig „generische“ Systeme sind

- **Verwaltung der Daten, die Daten beschreiben:**
  - Metadaten fallen **in allen DBS-Schichten** an
  - Synonyme: Metadatenverwaltung, DB-Katalog, Data-Dictionary-System, DD-System, ...

- **Transaktionsverwaltung**
  - **Realisierung der ACID-Eigenschaften** (Synchronisation, Logging/Recovery, Integritätssicherung)

- **Integration ins Schichtenmodell**
  - Lassen sich Metadaten- und TA-Verwaltung einer Schicht zuordnen?
  - Welche Schichten sind betroffen?
  - Hinweis: Was sind geeignete Granulate für Synchronisation usw. ?



Die ganze Wahrheit

- **Metadaten beschreiben vor allem semantische Aspekte der Daten**
  - Vollständigkeit und Genauigkeit dieser Beschreibung bestimmen „Eigenständigkeit“ und „Verhaltensflexibilität“ des Systems
  - hierarchische Anordnung der Beschreibungsaspekte zur **„semantischeren“** Spezifikation

- **(Versuch einer) Differenzierung der Metadaten**

Annotationen (Formate, Registrierungsdaten (z. B. bei Bildern))

Schemadaten (Strukturbezeichnungen, Integritätsbedingungen)

Beschreibungsdaten (Text, Schlüsselworte, Wissensrepräsentation)

Ontologien (Thesauri, Beziehungen zwischen den Begriffen/Konzepten)

zunehmende Semantikunterstützung

- **Meta-Metadaten (Metamodell)**
  - beschreiben die Metadaten (Modell)
  - Um die **syntaktische** Korrektheit eines Modells automatisch festzustellen, müssen das dazu zugehörige Metamodell und ggf. die verwendete Constraint-Sprache (OMG OCL[9]) bekannt sein
  - Kann man die **inhaltliche** Korrektheit eines Modells überhaupt ermitteln?

9. Object Constraint Language der Object Management Group

Die ganze Wahrheit (2)

- **Metadaten als „Daten über Daten“ ist nicht hinreichend genau:**

Data that describes the meaning and structure of business data, as well as how it is created, accessed, and used.
Devlin, B.: Data Warehouse: From Architecture to Implementation, Addison Wesley, 1997.

- **Korrektheit von Texten**
  - Syntax muss korrekt sein, bevor sich die Semantik erschließt
  - auf Wortebene, z. B. Masse oder Maße
  - auf Satzebene, aus der Beschreibung eines Pandas: "eats, shoots and leaves"

➥ **Was ein Komma alles anrichten kann!**

- **Aspekte der maschinelle Verarbeitung von Information**
  - Es kann niemals neue Information entstehen!
  - Es kann nur die vorhandene Information **sichtbar** (erkennbar, erfassbar) gemacht werden!

    ➥ Was bedeutet das bei „Datenbanken und Informationssystemen“? (Generierung von Berichten, Data Mining, Business Intelligence, ...)
  - Allerdings gibt es eine **entscheidende Voraussetzung**: Der Urtext muss fehlerfrei sein. Ist er es nicht – da reicht ein fehlendes Komma oder ein Tippfehler im Verb - schlägt die Satzanalyse fehl und heraus kommt großer Mist.
    Der Übersetzungscomputer zwischen den Ohren ist da einfach flexibler ;-)
  - Zu lang und komplex dürfen die Sätze weiterhin nicht sein, die magische Grenze liegt etwa bei 25 Wörtern. Einzelne unbekannte Wörter bringen die Satzanalyse ebenfalls zur Verzweiflung.
  - Sogar das **Übersetzen zusammengesetzter Wörter** klappt ganz gut, aber es wird nicht nur aus "Mathelehrer" "maths teacher", sondern auch aus "Oberammergau" "top bunting district".

Die ganze Wahrheit (3)

- **Test einer Online-Version eines Übersetzungsprogramms**

fruit flies like a banana -> Obstfliegen mögen eine Banane.

I'm impressed, but perhaps they taught it that sentence knowing we'd all test it with that! Let's try some more:

The big ship sailed on the ally-ally-o -> Das große Schiff segelte auf dem Verbündetenverbündeten-o.
John had his pants on back to front -> John hatte seine Hose zurück zu Vorderseite an.
He's so fine, wish he were mine, that handsome boy over there. -> Er ist so gut, wünschen, dass er Mine, dieser gut aussehende Junge dort, war.

Is the online version meant to be similar to the software you can buy?

- **Integration und Ableitung heterogener Information**
  - Semantic Web und Information Fusion wollen Information aus unbekannten Quellen durch Einsatz von Ontologien[10] zusammenführen und
  - nicht explizite Information durch automatische Schlussfolgerungen ableiten
  - Natural-language reasoning?

As we all know from elementary physics courses in high school:

(1) Power = Work / Time
(2) Knowledge = Power
(3) Time = Money
(4) Knowledgw = Work / Money
(5) Money = Work / Knowledge

Also each able manager is well aware, that

Applying (2), (3) to (1) we get

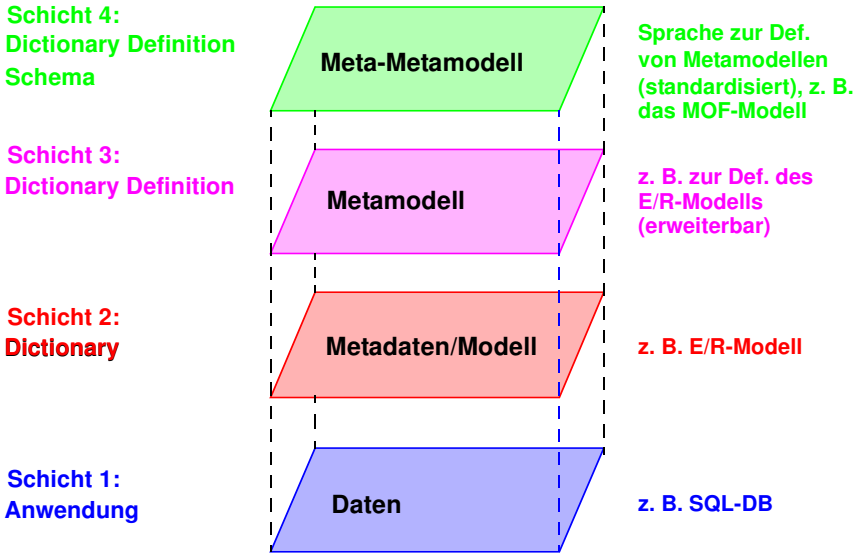
And finally

  - This law is known as the fundamental law on consultants pricing!

10. Ontologie ist ursprünglich eine philosophische Disziplin, neuerdings ein Modebegriff der Informatik. Nach Meyers Enzyklopädischem Lexikon ist Ontologie die „Lehre von dem Wesen und den Eigenschaften des Seienden”. In der Informatik ist sie die „formale Spezifikation eines bestimmten Gegenstandsbereichs in Form eines Begriffssystems”.

Meta, Meta-Meta, Meta-Meta-Meta

- **Wie spielen die Beschreibungsmodelle zusammen?**[11]
  - Um die syntaktische Korrektheit eines Modells durch ein Programm feststellen und um es generisch verarbeiten zu können, braucht man seine Beschreibung (sein Modell) auf der nächsthöheren Ebene
    - für die DB-Daten das DB-Schema
    - für ein UML-Modell (ER-Modell) das UML-Metamodell (das ER-Metamodell)
  - Sie sind sehr wichtig für automatische Modelltransformationen
    - Ziel: Programm- (System-) Generierungen aus Spezifikationen (= Modellen)
    - Ansätze: Generative SW-Entwicklung (engl. „model-driven architecture“)
- **Was bedeutet „Modellerweiterung“?**
- **Schichtenmodell des ANSI-IRDS oder OMG**[12]



11. CWM: Common Warehouse Metamodel, MOF: Meta Object Facility (Standard)
12. ANSI: American National Standards Institute, IRDS: Information Repository Definition Standard,

Schicht 1: Tabellen, Daten

Tabelle: KUNDE

KNR	NAME	ANSCHRIFT	...
1234	BAYER	KL ...	...
5678	SCHILCHER	SB...	...
6780	MITSCHANG	S ...	...

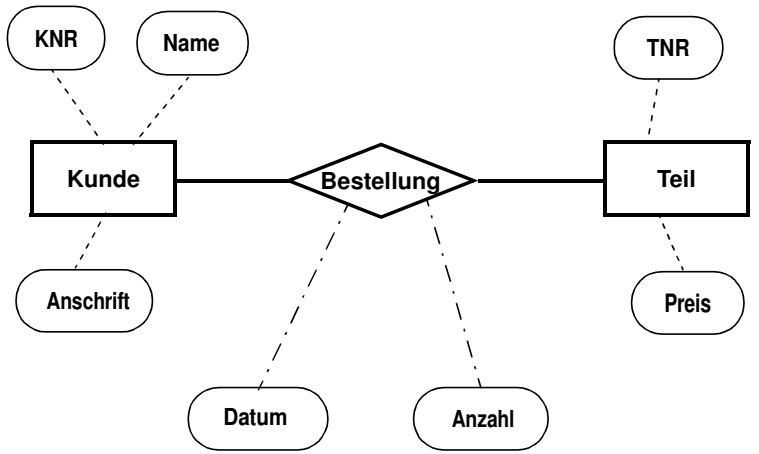
Tabelle: TEIL

TNR	PREIS	...
123 766	123.00	...
130 680	436.78	...
196 481	97.49	...

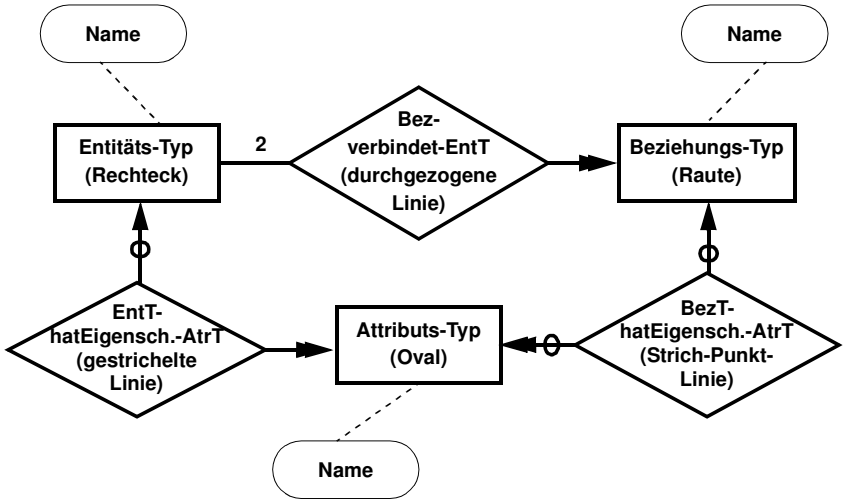
Tabelle: BESTELLUNG

KNR	TNR	ANZAHL	DATUM	...
1234	123 766	1000	1.2.00	...
1234	196 481	1500	2.7.99	...
5678	123 766	5000	4.9.98	...
5678	130 680	500	12.12.96	...
6780	130 680	3000	2.4.00	...
6780	196 481	3000	23.8.99	...

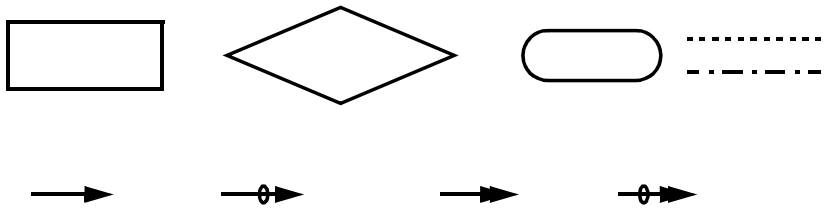
Schicht 2: ER-Modell



Schicht 3: Metamodell



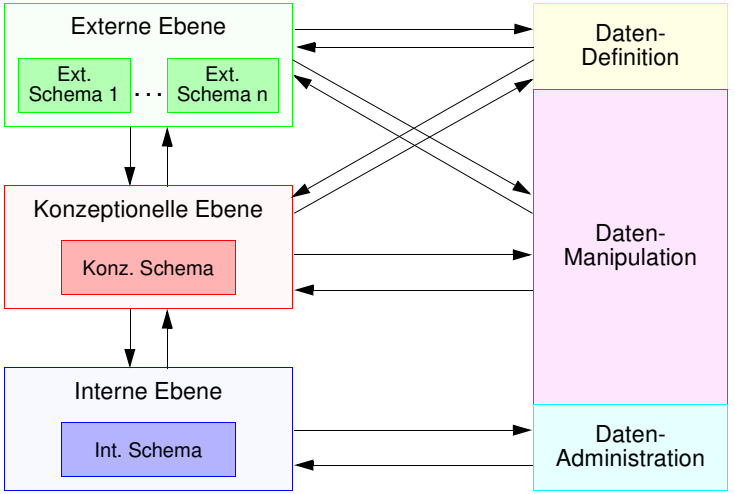
Schicht 4: Meta-Metamodell



Drei-Schema-Architektur[13] nach ANSI-SPARC

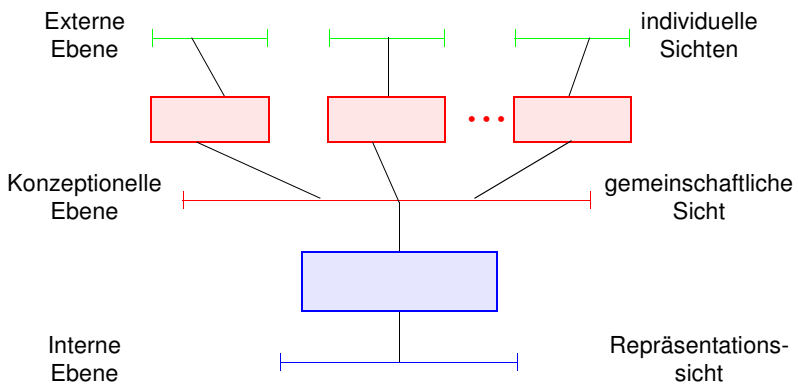
- **Verschiedene Betrachtungsweisen**
  - bisher: Realisierungssicht
    - ➥ Schichtenweiser Aufbau, Datenunabhängigkeit
  - jetzt: Benutzungssicht
    - ➥ Beschreibungsebenen, um aus dem „generischen“ ein „einsatzfähiges“ DBS (DBS-Installation) zu machen

- **3-Ebenen-Architektur nach ANSI/SPARC**



13. Tsichritzis, D. C., Klug, A.: The ANSI/X3/Sparc DBMS Framework Report of the Study Group on Database Management Systems, in: Information Systems 3:3, 1978, 173-191

Drei-Schema-Architektur (2)



Grobarchitektur für Schnittstellen nach ANSI/SPARC

- **Konzeptionelles Schema:**
  - (zeitvariante) globale Struktur; neutrale und redundanzfreie Beschreibung in der Sprache eines spezifischen Datenmodells
  - Beschreibungssprache: DDL (*Data Definition Language*)
- **Externes Schema:**
  - Definition von zugeschnittenen Sichten auf Teile des konzeptionellen Schemas für spezielle Anwendungen (Benutzer)
- **Internes Schema:**
  - legt physische Struktur der DB fest (physische Satzformate, Zugriffspfade etc.)
  - Beschreibungssprache: SSL (*Storage Structure Language*)
- **Gibt es noch weitere DB-Aspekte, die zu beschreiben sind?**

Drei-Schema-Architektur (3)

- **Stark vereinfachtes Beispiel für die Datenbeschreibung**

Extern (PL/1)

```
DCL   1 PERS,
        2 PNR CHAR(6),
        3 GEH FIXED BIN(31)
        4 NAME CHAR(20)
```

Extern (COBOL)

```
01   PERSC.
       02   PERSNR PIC X(6).
       02   ABTNR PIC X(4).
```

Konzeptionelles Schema

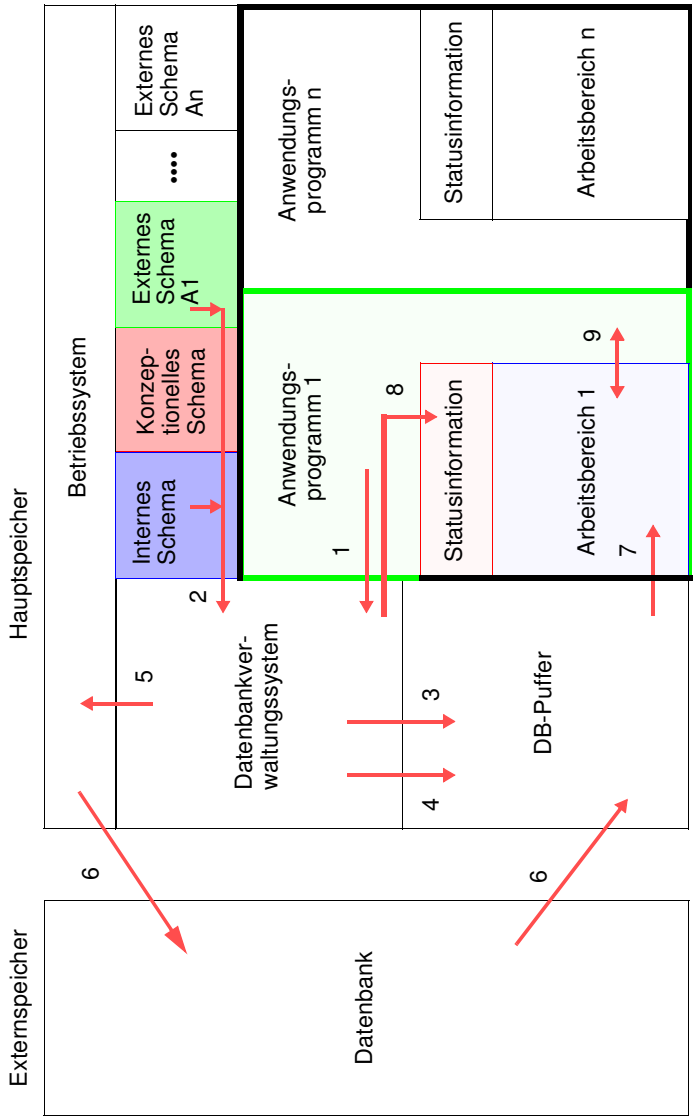
```
ANGESTELLTER
        ANG_NUMMER      CHARACTER   (6)
        NAME            CHARACTER   (20)
        ABT_NUMMER      CHARACTER   (4)
        GEHALT          NUMERIC     (5)
```

Internes Schema

```
SPEICHERUNG_PERS    LENGTH=40
        PREFIX      TYPE=BYTE(6), OFFSET=0
        PNR         TYPE=BYTE(6), OFFSET=6, INDEX=PNR
        NAME        TYPE=BYTE(20), OFFSET=12
        ANR         TYPE=BYTE(4), OFFSET=32
        GEHALT      TYPE=FULLWORD, OFFSET=36
```

- **Sichtenbildung durch das Externe Schema**
  - **Anpassung der Datentypen** an die der Wirtssprache (DBS ist „multi-lingual“)
  - **Zugriffsschutz:** Isolation von Attributen, Relationen, ...
  - **Reduktion der Komplexität:** nur die erforderlichen Daten sind für das Anwendungsprogramm sichtbar

Bearbeitung einer DB-Anweisung – dynamischer Ablauf



Bearbeitung einer DB-Anweisung – dynamischer Ablauf (2)

- **Beispiel-Schema:**

Konzeptionelles Schema: PERS (PNR, NAME, BERUF, ADR, ANR, GEHALT)
I5 . . . CHAR(50) . . .

Externes Schema: PERS' (PNR, BERUF, GEHALT, ANR)
PIC 9(5), PIC A(25), . . .

- **Interne Bearbeitungsschritte:**

1. SELECT * FROM PERS'
   WHERE PNR = '12345'
2. Vervollständigen der Verarbeitungsinformation aus Konzeptionellem und Internem Schema; Ermittlung der Seiten# (z. B. durch Hashing): P_{789}
3. Zugriff auf DB-Puffer: erfolgreich (weiter mit 7) oder
4. Zugriff auf die DB über DB-Pufferverwaltung/Betriebssystem
5. Durchführen des E/A-Auftrages
6. Ablegen der Seite im DB-Puffer

P_{789}

12345	Maier	Ing.	KL	K55	50000
• • •					

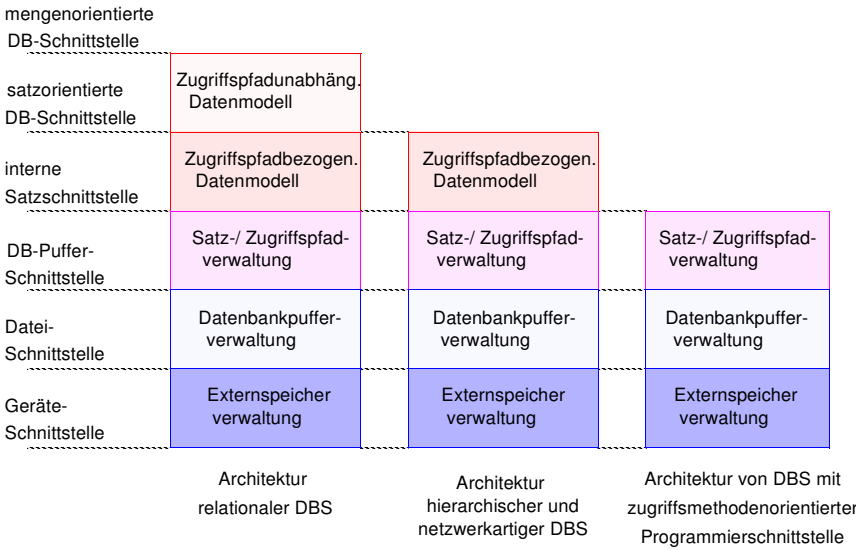
7. Übertragen nach Arbeitsbereich

12345	Ing.	50000	K55

8. Statusinformation: Return-Code, Cursor-Info
9. Manipulation mit Anweisungen der Programmiersprache

Vergleich von Programmierschnittstellen

- **Grobaufbau von DBS**



- **Vergleich der Typen von Programmierschnittstellen**

Anforderungen	**mengenorientiert**	**navigierend**	**zugriffsmethoden-orientiert**
Sprachebene	hoch	mittel	niedrig
Anzahl der DB-Anforderungen	1	1 (oder mehrere) pro aufgesuchtem Satz	1 (oder mehrere) pro aufgesuchtem Satz
aufgefundene Sätze	Satzmenge	1 Satz	1 Satz
Daten-verknüpfung	in der DB-Sprache	teils in der DB-Sprache, teils in der Wirtssprache	in der Wirts-sprache

Zusammenfassung

- **DBS-Charakteristika**
  - Zentralisierte Kontrolle über die operationalen Daten (Rolle des DBA)
  - Adäquate Schnittstellen (Datenmodell und DB-Sprache)
  - Zentrale Kontrolle der Datenintegrität und kontrollierter Mehrbenutzerbetrieb
  - Leistung und Skalierbarkeit
  - Hoher Grad an Datenunabhängigkeit

- **Beschreibungsmodelle für ein DBS (Beschreibung der Realisierung eines generischen DBMS)**
  - Schichtenmodelle
    ➥ Erklärungsmodelle für die statische Abbildungshierarchie
  - Rolle von Metadaten und Meta-Metadaten
  - Dynamisches Verhalten bei der Bearbeitung einer DB-Anfrage

- **Drei-Schema-Architektur (Spezifikation der Objekte eines konkreten DBS)**
  - Externes Schema zur Benutzerorientierung (Sichtenbildung)
  - Konzeptionelles Schema als logische und neutrale DB-Beschreibung
  - Internes Schema als Beschreibung der physischen DB-Aspekte

- **Programmierschnittstellen (APIs, DB-Sprachen) (siehe Schichtenmodell)**
  - mengenorientierte DB-Schnittstelle
    ➥ relationale DBS
  - satzorientierte DB-Schnittstelle
    ➥ hierarchische und netzwerkartige DBS
  - interne Satzschnittstelle (zugriffsmethodenorientierte API)
    ➥ „DMS“