



Kapitel 5 – Datenbankentwurfprozess

Vorlesung Datenbanken

Dr. Kai Höfig



Kapitel 3: Datenbankentwurf

3.1 Entwurfsaufgaben

3.2 Phasen des Datenbankentwurfs



Ziel des Datenbankentwurfs

- ◆ Je besser der Entwurf der Datenhaltung, desto länger sind die Daten über Jahre und Jahrzehnte verwendbar.
 - Änderung der Anwendungen/Anwendungsszenarien
 - Änderung der nicht-funktionalen Anforderungen z.B. an Performanz, Verfügbarkeit oder Verteilung
 - Besondere Bedeutung von Redundanzvermeidung um Anomalien zu vermeiden und Speicherplatz zu sparen.

- ◆ Neben der Frage WIE Daten möglichst sinnvoll gespeichert werden stellt sich auch die Frage WAS gespeichert werden soll
 - Zu wenig Informationen decken möglicherweise nicht den zukünftigen Informationsbedarf eines Unternehmens
 - Alle anfallenden Daten zu speichern ist vielleicht sogar interessant (Big data, machine learning), aber teuer.



♦ Datenbankentwurf

- Abfolge von Entwurfsdokumenten (**Modellierungen**)
- Beginnend bei abstrakter, anwendungsnaher Beschreibungsebene
- Endend in tatsächlicher Realisierung einer Datenbank
- **Entwurfsschritt**: von einem Modell zum nächsten (manuell, automatisch oder semi-automatisch)

♦ Wichtigste Eigenschaften von **Entwurfsschritten**

1) **Informationserhalt**

Es können alle Daten im neuen Modell gespeichert werden, die auch im vorherigen gespeichert werden konnten

2) **Konsistenzerhaltung**

Regeln und Einschränkungen im vorherigen Modell können auch im neuen gewährleistet werden

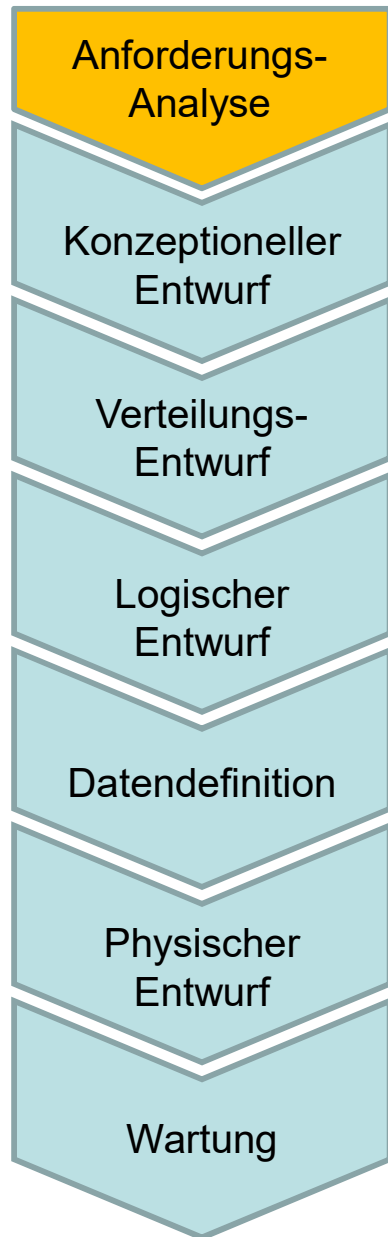


Gütekriterien für Entwurfsdokumente

- ◆ Redundanzfreiheit
- ◆ Vollständigkeit bezüglich Anforderungsanalyse
- ◆ Konsistenz des Beschreibungsdokuments
- ◆ Ausdruckstärke, Verständlichkeit des benutzten Formalismus (formale Semantik der Beschreibungs-Konstrukte)
- ◆ Lesbarkeit der Dokumente
- ◆ Weitere Qualitätseigenschaften: Erweiterbarkeit, Modularisierung, Wiederverwendbarkeit, Werkzeugunterstützung etc.



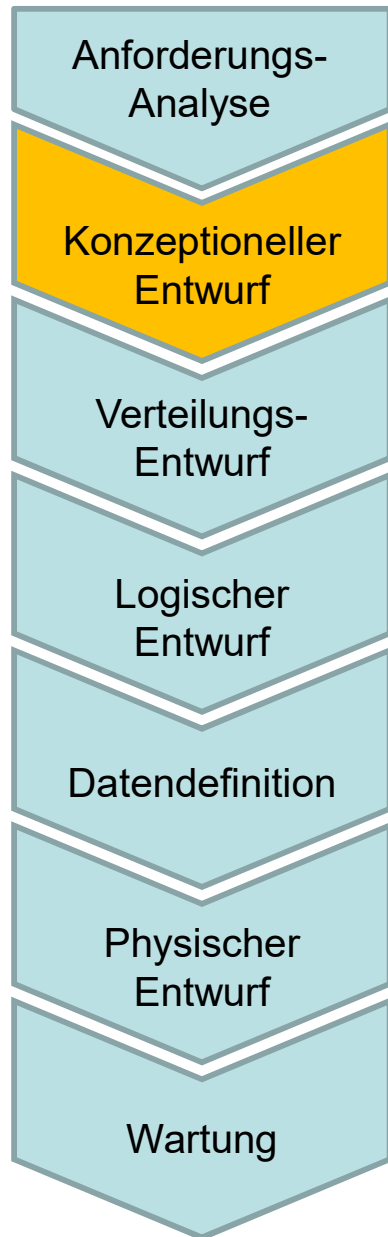
Phasen des Datenbankentwurfs



- ◆ **Ziel:** Sammeln und Analysieren der Anforderungen („Requirements“) an das zu realisierende Datenbanksystem
- ◆ **Vorgehensweise:** Sammlung des Informationsbedarfs in den Fachabteilungen
- ◆ **Ergebnis:**
 - informale Beschreibung (Texte, tabellarische Aufstellungen, Formblätter, usw.) des Fachproblems
 - Trennen der Information über Daten (Datenanalyse) von den Information über Funktionen (Funktionsanalyse)
- ◆ In dieser Vorlesung nur gestreift, typische Methoden siehe Vorlesung **Software Engineering**



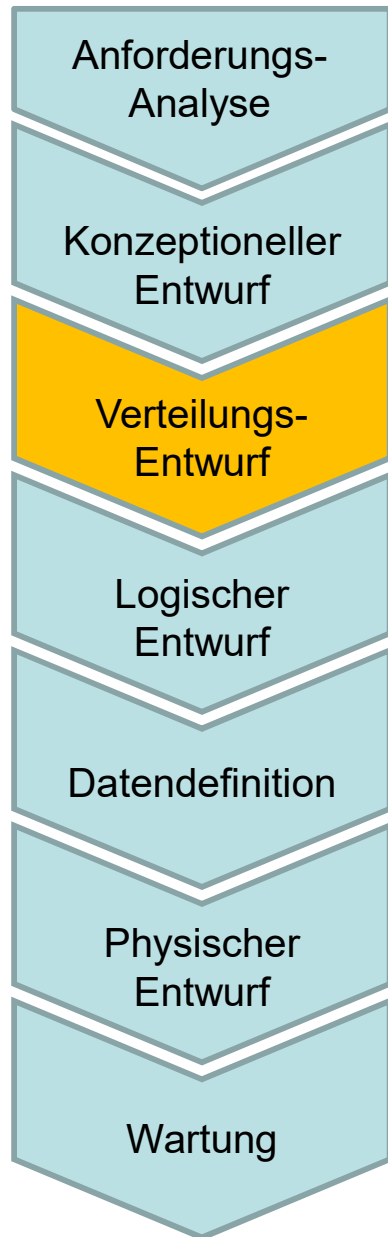
Phasen des Datenbankentwurfs



- ◆ **Ziel:** Erste formale Beschreibung des Fachproblems, *unabhängig* vom später zu verwendenden System
- ◆ **Sprachmittel:** abstraktes (semantisches) Datenmodell
- ◆ **Vorgehensweise:**
 - Modellierung von Sichten z.B. für verschiedene Fachabteilungen
 - Analyse der vorliegenden Sichten untereinander in Bezug auf Konflikte, wie z.B.
 - Namenskonflikte: Homonyme / Synonyme
 - Typkonflikte: verschiedene Strukturen für das gleiche Element
 - Wertebereichskonflikte: verschiedene Wertebereiche für ein Element
 - Bedingungskonflikte, z.B. verschiedene Schlüssel für ein Element
 - Strukturkonflikte: gleicher Sachverhalt durch unterschiedliche Konstrukte ausgedrückt
 - Integration der Sichten in ein konsistentes und widerspruchsfreies Gesamtschema
- ◆ **Ergebnis:** konzeptionelles Datenmodell, typischerweise ein ER- oder UML-Diagramm
- ◆ **Integraler Bestandteil dieser Vorlesung**



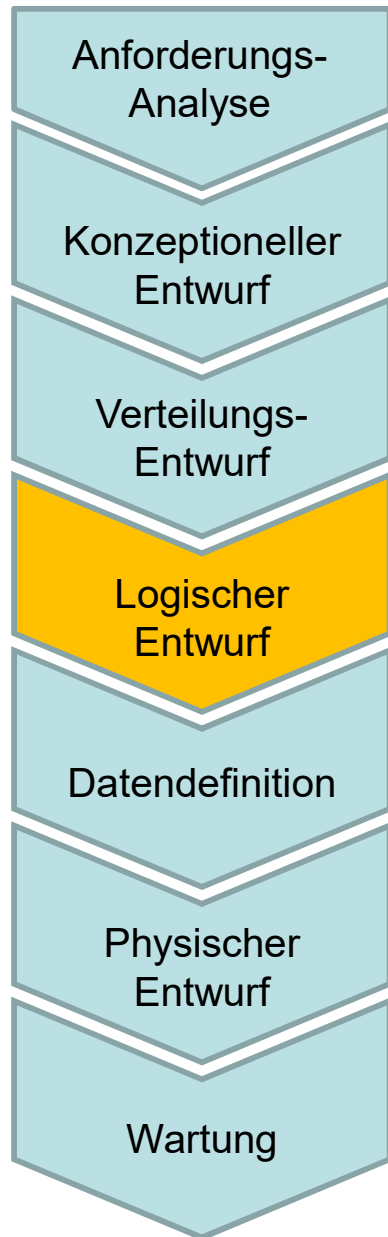
Phasen des Datenbankentwurfs



- ◆ Optional
- ◆ Sollen Daten auf mehreren Rechnern verteilt vorliegen, muss Art und Weise der **verteilten Speicherung**, festgelegt werden
- ◆ U.U. auch erst im Rahmen des physischen Entwurfs
- ◆ In dieser VL nicht weiter betrachtet, siehe Vorlesung **Datawarehouse**



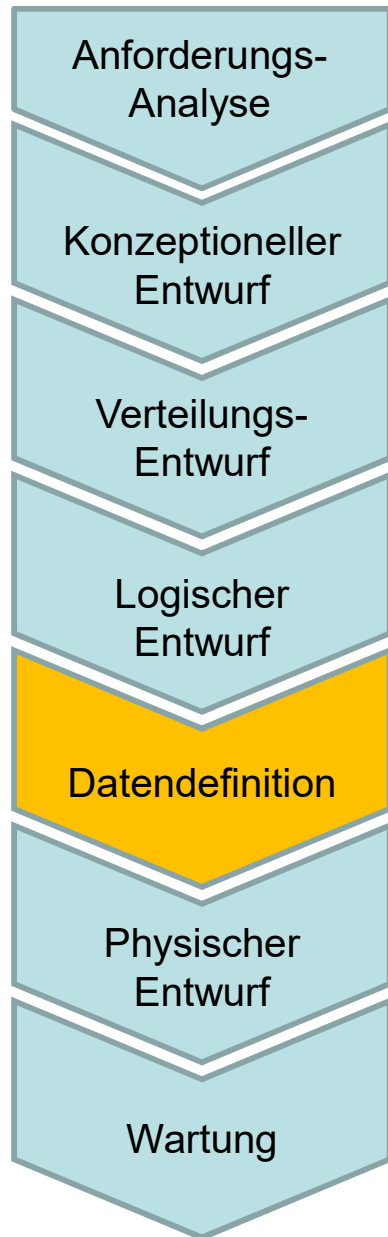
Phasen des Datenbankentwurfs



- ♦ **Vorgehensweise:**
 - (automatische) Transformation des konzeptionellen Schemas
z.B. ER → relationales Modell
 - Verbesserung des relationalen Schemas anhand von Gütekriterien
(Stichwort: Normalisierung, siehe später)
Entwurfsziele: Redundanzvermeidung, . . .
- ♦ **Sprachmittel:** Datenmodell des ausgewählten Realisierungs-DBMS z.B. relationales Modell
- ♦ **Ergebnis:** logisches Schema, z.B. Sammlung von Relationenschemata
- ♦ **Integraler Bestandteil dieser Vorlesung**



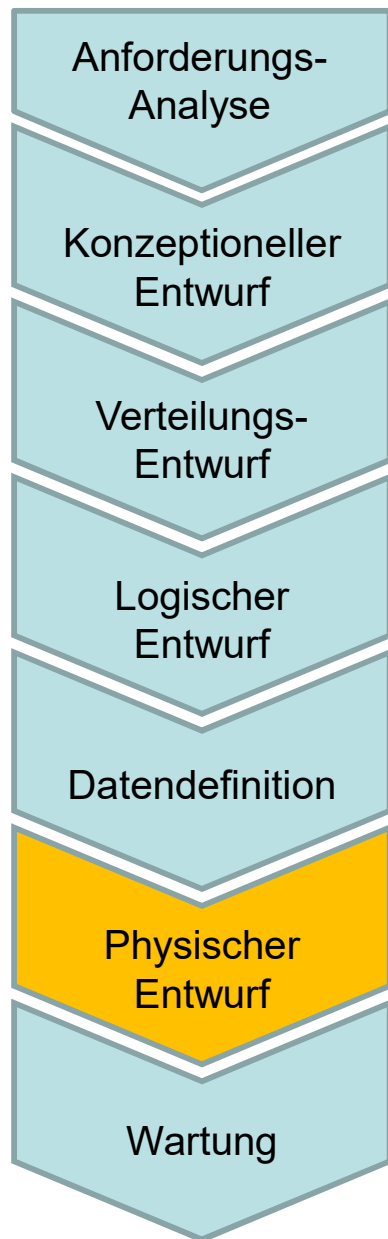
Phasen des Datenbankentwurfs



- ◆ **Ziel:** Umsetzung des logischen Schemas in ein konkretes Schema
- ◆ **Sprachmittel:** DDL und DML eines DBMS z.B. Oracle, IBM DB2, Microsoft SQL Server
 - Datenbankdeklaration in der DDL des DBMS
 - Realisierung der Integritätssicherung
 - Definition der Benutzersichten
- **Integraler Bestandteil dieser Vorlesung**



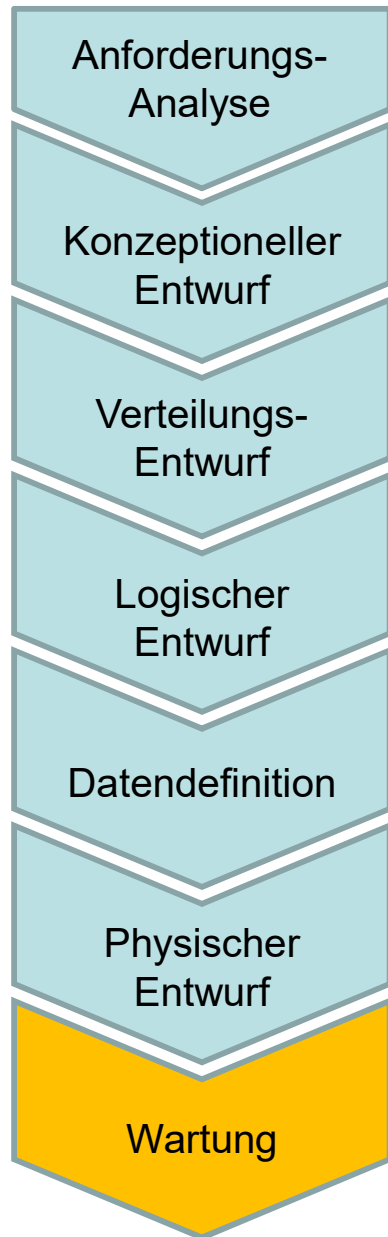
Phasen des Datenbankentwurfs



- ◆ **Ziel:** Ergänzen des physischen Entwurfs um Zugriffsunterstützung bzgl. Effizienzverbesserung, z.B. Definition von Indexen
- ◆ **Index**
 - Zugriffspfad: Datenstruktur für zusätzlichen, schlüsselbasierten Zugriff auf Tupel (<Schlüsselattributwert, Tupeladresse>)
 - meist als B*-Baum realisiert
- ◆ **Sprachmittel:** Speicherstruktursprache (z.B. als Teil von SQL)
- ◆ **Notwendigkeit für Zugriffspfade**
 - Beispiel: Tabelle mit 100 GB Daten, Festplattentransferrate ca. 50 MB/s
 - Operation: Suchen eines Tupels (Selektion)
 - Implementierung: sequentielles Durchsuchen
- ◆ **Grundlagen als Teil der Vorlesung**



Phasen des Datenbankentwurfs



- ◆ Typische Herausforderungen der Wartungsphase:
 - Weiteren Optimierung der physischen Ebene
 - Anpassung an neue Anforderungen und Systemplattformen
 - Portierung auf neue Datenbankmanagementsysteme
 - etc.
- ◆ In dieser VL nicht weiter betrachtet



Phasenbegleitende Methoden

Phasenbegleitende Methoden zur Sicherstellung der Anforderungen an den Entwurf (v.a. Informations- und Konsistenzerhaltung)

- ♦ **Verifikation**: Der formale Beweis etwa von Schemaeigenschaften
- ♦ **Prototyping**: beispielhaftes Arbeiten mit der Datenbank vor der endgültigen Implementierung
- ♦ **Validation mit Testdaten**: Überprüfung der Richtigkeit des Entwurfs anhand von realen oder künstlichen Testdaten



- ◆ Datenbankentwurf als **Abfolge von Entwurfsschritten** (Modellierungen)
- ◆ Start bei den abstrakten Anforderungen der Fachabteilungen, hinarbeitend auf ein konkretes, implementiertes DBS
- ◆ Phasen des Datenbankentwurfs
 - Anforderungs-Analyse
 - **Konzeptioneller Entwurf**
 - Verteilungs-Entwurf
 - **Logischer Entwurf**
 - **Datendefinition**
 - Physischer Entwurf
 - Wartung