

---

# **Data Warehouse Schema Design**

Jens Lechtenbörger

Institut für Wirtschaftsinformatik

Westfälische Wilhelms-Universität Münster

<http://dbms.uni-muenster.de>

---

# „Wir“

---

- Bodo Hüsemann
  - Dominique Laurent
  - Nicolas Spyrtos
  - Gottfried Vossen
-

# Motivation

---

- Datenlager: Integrierte Datenbank für Entscheidungsunterstützung
- Datenbankentwurf gut verstanden
  - Methodisches Vorgehen (Phasenmodelle)
  - Normalformen als Gütekriterien für Schemata
- Schemaentwurf für Datenlager
  - Ad-Hoc-Vorgehen
  - Star- oder Snowflake-Schemata “relevant”
  - Methodisches Vorgehen (aus Datenbanksicht) wenig untersucht
- Ziel: Entwurfsprozess für Datenlagerschemata
  - Anlehnung an traditionellen Datenbankentwurf
  - Identifikation und Durchsetzung “guter” Schemaeigenschaften
  - Integration und Erweiterung unabhängiger Methoden

# Überblick

---

Datenlagerentwurf ist *Datenbankentwurf*

- Anforderungsanalyse und -spezifikation
- Konzeptioneller Entwurf
  - Ziel: Summierbarkeit, Erfüllung mehrdimensionaler Normalformen
  - Zentrales Hilfsmittel: *Funktionale Abhängigkeiten*
- Logischer Entwurf
  - Ziel: Unabhängigkeitseigenschaften
  - Zentrales Hilfsmittel: *Komplemente*
- (physischer Entwurf)

# Gliederung

---

- Motivation und Überblick
  - **Zugrunde liegende Arbeiten**
  - Konzeptioneller Entwurf
  - Logischer Entwurf
  - Zusammenfassung
-

# Auswahl zugrunde liegender Arbeiten

---

**Datenbankentwurf** Batini, Ceri, Navathe, 1992

**Datenlagerentwurf** Golfarelli, Rizzi, 1998

**Summierbarkeit** Lenz, Shoshani, 1997

**Mehrdimensionale Normalformen** Lehner, Albrecht, Wedekind, 1998

**Generalisierungen/Spezialisierungen** Smith, Smith, 1977

**Selbstwartbarkeit** Gupta, Jagadish, Mumick, 1994

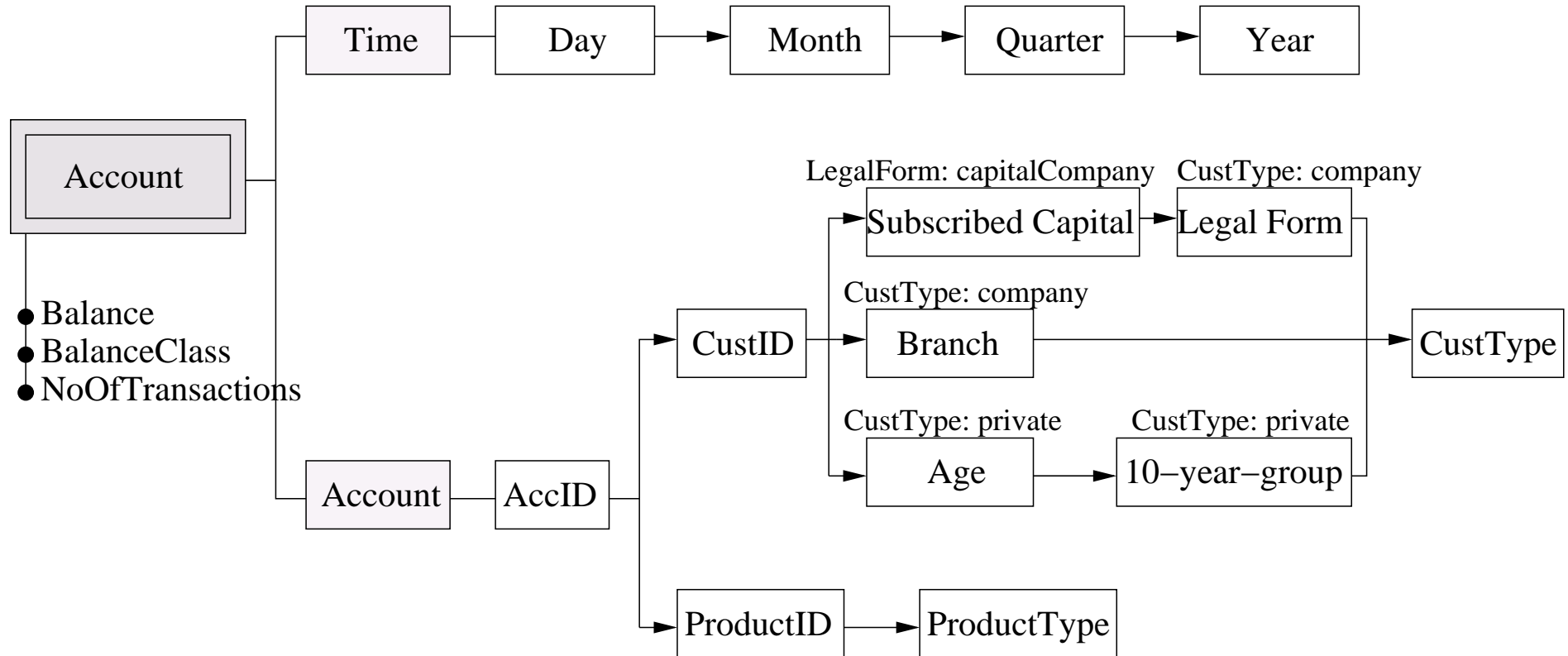
**Komplemente** Bancilhon, Spyrtos, 1981

# Gliederung

---

- Motivation und Überblick
  - Zugrunde liegende Arbeiten
  - **Konzeptioneller Entwurf**
  - Logischer Entwurf
  - Zusammenfassung
-

# Konzeptioneller Entwurf (1/2)



Semantik von Faktenschema  $F$ :

Relation über Universum mit funktionalen Abhängigkeiten  $FD_F$

# Konzeptioneller Entwurf (2/2)

---

## Implikationen der dritten mehrdimensionalen Normalform (3MNF)

- „Vernünftige“ Beziehung zwischen Datenquellen und Datenlager
  - Treue Abbildung der Anwendungswelt
  - Vollständigkeit des Datenlagers bezüglich der Anwendungswelt
  - Vermeidung von Redundanzen
- Kontrollierte Anwendung optionaler Dimensionsebenen
  - Kontext-sensitive Summierbarkeit
  - Vermeidung widersprüchlicher Anfragen
  - Effizienter physischer Entwurf (Vermeidung von Nullwerten)

## Konzeptioneller Entwurf (2/2)

---

### Implikationen der dritten mehrdimensionalen Normalform (3MNF)

- „Vernünftige“ Beziehung zwischen Datenquellen und Datenlager
  - Treue Abbildung der Anwendungswelt
  - Vollständigkeit des Datenlagers bezüglich der Anwendungswelt
  - Vermeidung von Redundanzen
- Kontrollierte Anwendung optionaler Dimensionsebenen
  - Kontext-sensitive Summierbarkeit
  - Vermeidung widersprüchlicher Anfragen
  - Effizienter physischer Entwurf (Vermeidung von Nullwerten)

### Konzeptioneller Entwurf

- Basierend auf Analyse funktionaler Abhängigkeiten
- Garantiert 3MNF

# Gliederung

---

- Motivation und Überblick
  - Zugrunde liegende Arbeiten
  - Konzeptioneller Entwurf
  - **Logischer Entwurf**
  - Zusammenfassung
-

# Logischer Entwurf (1/2)

---

**Ziel:** Menge materialisierter Sichten, die Faktenschemata implementiert

**Randbedingung:** Operationale Datenquellen nur eingeschränkt abfragbar

**Konsequenz:** Keine Anfragen an Datenquellen während Wartung und Betrieb

**Formal:** Unabhängigkeit

# Logischer Entwurf (1/2)

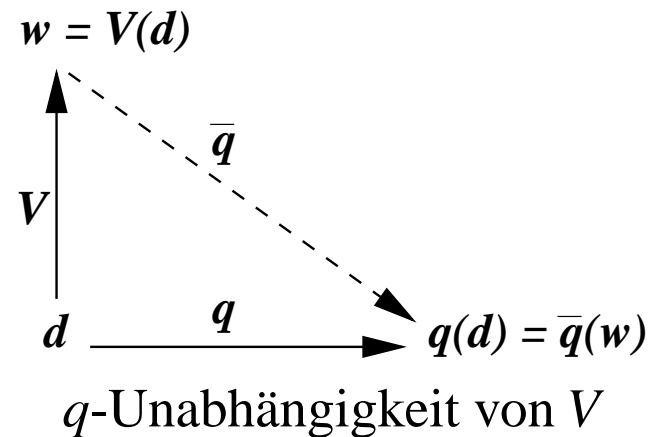
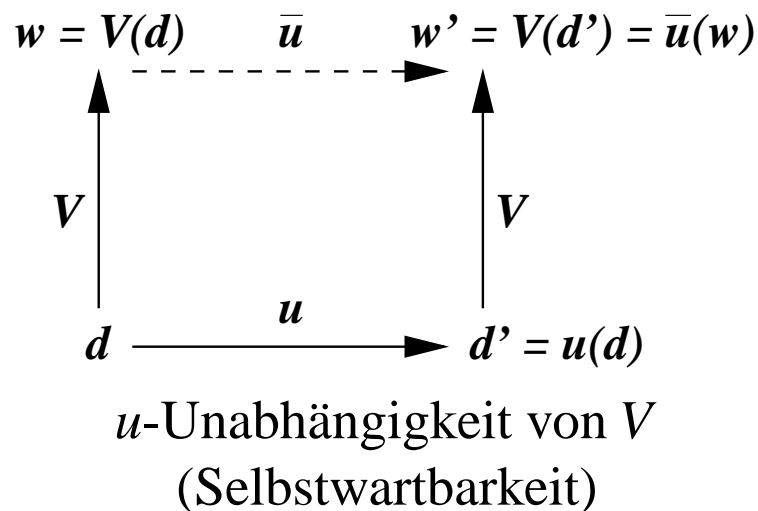
**Ziel:** Menge materialisierter Sichten, die Faktenschemata implementiert

**Randbedingung:** Operationale Datenquellen nur eingeschränkt abfragbar

**Konsequenz:** Keine Anfragen an Datenquellen während Wartung und Betrieb

**Formal:** Unabhängigkeit

$V$  Datenlager über Datenquellen  $D$ ,  $q$  Anfrage und  $u$  Änderung über  $D$ .



# Logischer Entwurf (2/2)

---

## Lemma:

1. Projektion  $\pi_Z(R)$ , die keinen Schlüssel von  $R$  erhält, ist *nicht* selbstwartbar.
2. Verbund verschiedener Relationen ist *nicht* selbstwartbar.

# Logischer Entwurf (2/2)

---

## Lemma:

1. Projektion  $\pi_Z(R)$ , die keinen Schlüssel von  $R$  erhält, ist *nicht* selbstwartbar.
2. Verbund verschiedener Relationen ist *nicht* selbstwartbar.

## Beobachtungen

- Selbstwartbarkeit durch Hinzufügen zusätzlicher Sichten möglich.
- Wünschenswert ist „kleine“ Menge zusätzlicher Sichten.

# Logischer Entwurf (2/2)

---

## Lemma:

1. Projektion  $\pi_Z(R)$ , die keinen Schlüssel von  $R$  erhält, ist *nicht* selbstwartbar.
2. Verbund verschiedener Relationen ist *nicht* selbstwartbar.

## Beobachtungen

- Selbstwartbarkeit durch Hinzufügen zusätzlicher Sichten möglich.
- Wünschenswert ist „kleine“ Menge zusätzlicher Sichten.

## Resultate

- Komplemente erlauben Berechnung „kleiner“ Menge.
- Effiziente Berechnung von Komplementen für UPSJR-Sichten.
- Selbstwartbarkeit basierend auf Komplementen algorithmisch durchsetzbar.

# Gliederung

---

- Motivation und Überblick
  - Zugrunde liegende Arbeiten
  - Konzeptioneller Entwurf
  - Logischer Entwurf
  - **Zusammenfassung**
-

# Zusammenfassung

---

Datenlagerentwurf ist *Datenbankentwurf*

- Anforderungsanalyse und -spezifikation
- konzeptioneller Entwurf
  - Mehrdimensionale Normalformen
  - *Funktionale Abhängigkeiten* als zentrales Werkzeug
- logischer Entwurf
  - Unabhängigkeitseigenschaften
  - *Komplemente* als zentrales Werkzeug

<http://dbms.uni-muenster.de>

---

J. Lechtenbörger, *Data Warehouse Schema Design*, DISDBIS 79, Infix, 2001

<http://dbms.uni-muenster.de>

---

J. Lechtenbörger, *Data Warehouse Schema Design*, DISDBIS 79, Infix, 2001

**Vielen Dank für Ihre Aufmerksamkeit!**