

# 3. Aufgabenblatt - - Auswertung -

**Patrick Schäfer**

Berlin, 18. Dezember 2017

[patrick.schaefer@hu-berlin.de](mailto:patrick.schaefer@hu-berlin.de)

# Szenario

- Wir bauen ein Data Warehouse für einen nationalen Autoreparaturkonzern, der Informationen über die in seinen Werkstätten durchgeführten Fahrzeugreparaturen sammeln möchte.
- Die Reparaturen werden von Werkstätten durchgeführt. Diese haben einen Typ (Händler oder Werkstatt) und liegen in einem Bundesland; außerdem hat jede Werkstatt eine Größe, die sich durch die Zahl verfügbarer Mechaniker ausdrückt.
- Eine Reparatur wird durchgeführt an einem Fahrzeug, das ein bestimmtes Modell eines Herstellers ist und in einem bestimmten Jahr hergestellt wurde.
- Die Reparatur hat einen Anfangstermin und einen Endtermin.
- Das reparierte Auto gehört einem Kunden, der einen Vor- und Nachnamen hat, sowie eine Adresse (Straße) und ein Bundesland.
- Jede Reparatur verursacht Ersatzteilkosten und Personalkosten der beteiligten Mitarbeiter
  - Beides zusammen ergibt die Gesamtkosten

# Kosten

- Personalkosten
  - An einer konkreten Reparatur X arbeiten mehrere Personen einer Werkstatt
  - Diese verursachen individuelle Kosten, die modelliert werden müssen, z.B.:
    - X, Müller, 1000
    - X, Meier, 700
    - ...
- Pro Person gibt es pro Reparaturvorgang nur eine Kostenzahl
- Die Ersatzteilkosten sind dagegen nur eine Zahl pro Reparatur
- Die Gesamtkosten einer Reparatur dürfen nicht explizit gespeichert werden, sondern setzen sich aus den Ersatzteil- und den Personalkosten zusammen
- Ebenso wenig darf die Anzahl der an einer Reparatur beteiligten (und Personalkosten verursachenden) Mitarbeiter explizit gespeichert werden

# Anmerkungen

- Keine Fremdschlüssel auf Faktentabelle vorgesehen in M/ER
- Keine M:N Beziehung modellieren zwischen Faktentabelle und Dimensionstabelle
- Fakten sind aggregierbar
- Alle Fakten in Würfel speichern
- Einheitliche Namensgebung (M/ER, Dimensionen, Star/Schema, ROLAP)
- Hierarchie: Feingranular zu Grobgranular

# 12 Abgaben

## Gruppe

---

Blauer Vogel

CJLdwhdm

dsh

DWHMiMoJoin2017

Gruppe 1

Join Venture

MondayMorning

SMH

SpeedyGonzales

TeamHA9876

Wwgds

GrundDan

---

# Anmerkungen im Detail

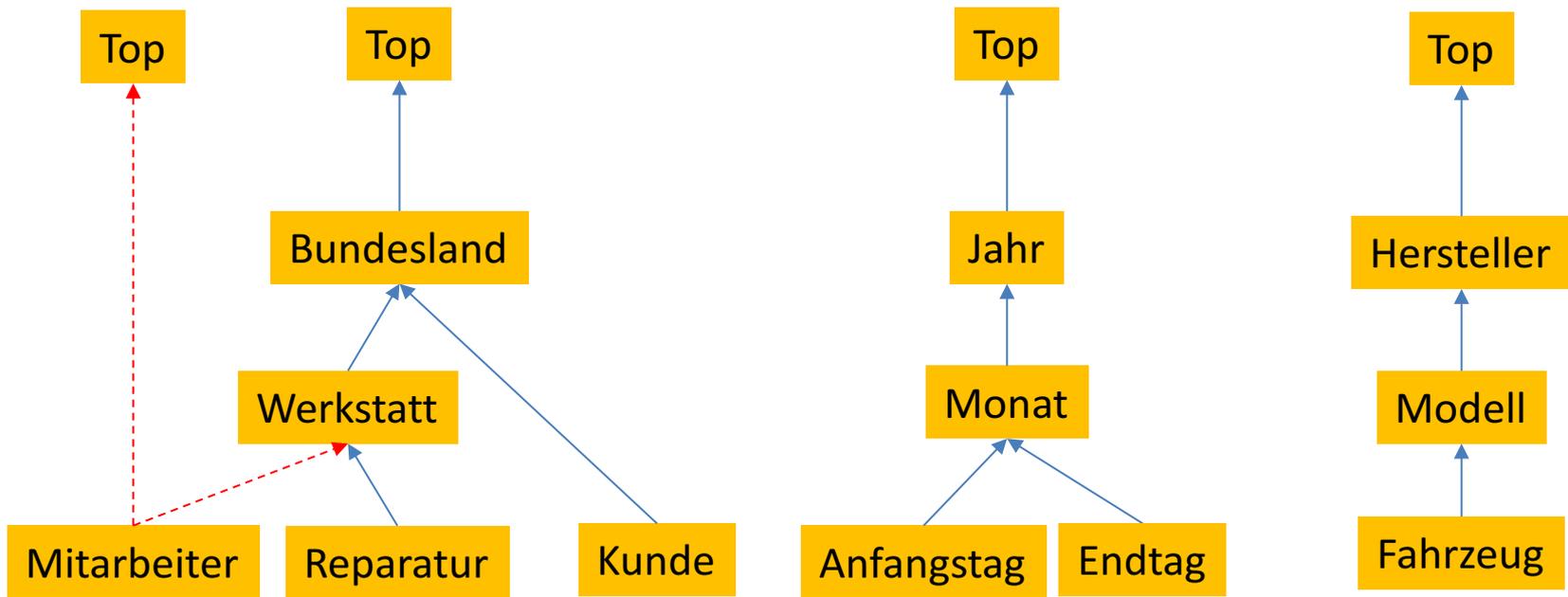
Menüs durchsuchen (Wahlt)									
fx									
	A	B	C	D	E	F	G	H	
Übung 3	Gruppe	Abgegeben	Kommentar	Dimensionen	MER	Star Schema	Snowflake Schema		
2	Blauer Vogel	OracleJ	ok	1) Folie 3 "Die Gesamtkosten einer Reparatur dürfen nicht explizit gespeichert werden, sondern setzen sich aus den Ersatzteil- und den Personalkosten zusammen" 2) PErsonalkosten sollten eigentlich nicht als Ganzes modelliert werden	1) Fakt -> Kosten (Ersatzteil + Personal). Also etwas aggregierbares 2) Mitarbeiter fehlen als Dimension 3) Anfangs-TAG und End-Tag als Zeitdimension 4) Straße ist eher Attribut von Kunde 5) Fahrzeug fehlt als Dimension. Momentan sind alle Modelle im gleichen Baujahr?	1) Aufgabenstellung: Einzelkosten für Mitarbeiter fehlen	1) Mitarbeiter und Einzelkosten fehlen 2) Einheitliche Namensgebung verwenden zwischen MER und Star/Snowflake	1) Siehe Star Schema	ok
4	CJLdwhm	OracleE	ok		1) Kosten sind die Fakten 2) Adresse ist besser Attribut als Hierarchie 3) Anfangs-TAG und End-Tag als Zeitdimension	1) Personalkosten gehören in Reparatur	1) 1:N Beziehung wurde modelliert zwischen Reparatur und Mitarbeiter. Nicht erlaubt im Star Schema. Reparatur muss alle Fakten enthalten und Fremdschlüssel auf Dimensionstabellen enthalten. 2) Einheitliche Bezeichner verwenden!	1) Folgefehler mit 1:N Beziehung	GI Ta N
5	dsh	ORadeF	ok	1) Quell Kurze Beschreibung wäre schön, was Kostentafel ist	1) Fahrzeugmodell fehlt 2) 1:1 Beziehung zwischen Reparatur, Anfangstag, Endtag und Fahrzeug. Daher überflüssig	1) Kostentafel = Mitarbeiter und Ersatzteilkosten?	1) Einheitliche Namen verwenden zwischen MER und Star und Snowflake Schema	1) Mitarbeiter und dessen Kosten fehlen	
6	DWHMMJoin2017	OracleB	ok	1) Fakten sollten aggregierbar sein 2) N:M Beziehung wurde modelliert	1) Reparatur ist kein Fakt (nicht aggregierbar) 2) Straße ist besser Attribut als Hierarchie 3) (Fahrzeug-Modell fehlt 4) Star-Enddatum->Monat -> Jahr	1) N:M Beziehung modelliert. Die sind nicht erlaubtvorgesehen, auch nicht durch Einfügen eines weiteren (Pseudo-Cubes	1) N:M Beziehung wurde modelliert zwischen Reparatur und Mitarbeiter. Nicht erlaubt im Star Schema. Reparatur muss alle Fakten enthalten und Fremdschlüssel auf Dimensionstabellen enthalten.	1) Siehe Star Schema	ok
7	Gruppe 1	OracleC	ok		1) Fakten sind aggregierbar -> Startzeit, Endzeit, Mechaniker sind keine Fakten	1) Fakten kommen immer in den Würfeln! Ihr habt Personalkosten und Ersatzteilkosten als Fakt identifiziert... 2) Anfangs-Tag und End-Tag sind die	1) Personalkosten in den Würfeln	1) siehe Star Schema	ok
8	Join Venture	OracleH					modelliert Mitarbeiter. Nicht reparatur muss Fremdschlüssel (halten. Sonst ist sinnvoll ersatzteilkosten)	1) Folgefehler mit N:M Beziehung	ok
9	MondayMorning	OracleG					ro Bundesland nicht über 9	1) siehe Star Schema	ok
10	SMH	OracleI	ok	1) Variable Kosten pro Mitarbeiter und Reparatur? 2) Kosten fehlen in der Fakten-Tabelle (das worüber ein Aggregat gebildet wird)	1) Ersatzteilkosten sind ein Fakt über den Anfragen (Aggregat) gestellt werden 4) Ersatzteilkosten sind ein Fakt über den Anfragen (Aggregat) gestellt werden	müssen (sonst mögliche Inkonsistenzen) 3) Ende und Anfang sind Ende-tag und Anfang-tag	1) SQL: Ersatzteilkosten pro Bundesland sehr kompliziert 2) HStar: insufficient für Anfrage, da immer gefiltert werden muss. Effizienter wären getrennte Tabellen	1) siehe Star Schema	ok
11	SpeedyGonzales	OracleA	ok		1) Ersatzteilkosten als Fakt speichern 2) Können Mitarbeiter in verschiedenen Werkstätten arbeiten? 3) Kunden können mehrere Autos haben! Daher falsche Hierarchie/Granularität: bei each grob (Kunde) zu fein (Auto). Sollte aber sein kein (Auto) zu grob (Kunde)	1) Es muss jetzt immer über Reparatur gruppiert werden, um Ersatzteilkosten zu bekommen 2) Variable Kosten pro Mitarbeiter und Reparatur?	1) SQL: Ersatzteilkosten pro Bundesland sehr kompliziert	1) siehe Star Schema	ok
12	TeamHA9876	OracleD	ok	1) Folie 3 "Die Gesamtkosten einer Reparatur dürfen nicht explizit gespeichert werden, sondern setzen sich aus den Ersatzteil- und den Personalkosten zusammen" 2) Im MER sind keine 1:N Beziehungen zwischen Faktentabelle und Dimension erlaubt. Aber mehrere Mitarbeiter können an einer Reparatur arbeiten. Bei auch nur einer (siehe Star und Snowflake Schema). 3) PErsonalkosten sollten eigentlich nicht als Ganzes modelliert werden	1) Dimension Fahrzeug -> Model -> ... (Momentan sind alle Modelle im gleichen Baujahr?) 2) Fakt Gesamtkosten nicht erlaubt 3) Modelliert werden sollen alle an einer Reparatur beteiligten Mitarbeiter. Fakt Personalkosten (als Summe) nicht erlaubt.	1) 1:N Beziehung zwischen Reparatur und Mitarbeiter nicht erlaubt im MER Modell 2) Alle Modelle immer im gleichen Baujahr?	1) individuelle Mitarbeiterkosten? Müller 1000, Meier 700, ...? 2) Reparatur hat so nur einen Mitarbeiter!	1) siehe Star Schema	ok
13	wwgds	OracleK	ok	1) Variable Kosten pro Mitarbeiter und Reparatur? 2) Viele ID-Felder in Star Schema fehlen 3) Anfangs und Enddatum als Dimension?	1) Dimension "Datum": Anfangs oder Enddatum modelliert? 2) Dimension "Ort" für Mitarbeiter oder Werkstatt? 3) Können Mitarbeiter in verschiedenen Werkstätten arbeiten?	1) Wie werden variable Personalkosten modelliert? So sind nur fixe Personalkosten zu modellieren 2) Besser: Werkstattgröße direkt speichern	1) ID-Felder fehlen (siehe Snowflake Schema): was wenn z.B. gleicher Modellname bei mehreren Herstellern? 2) Folgefehler: Reparaturanfang-ende	1) Wieso sind hier die ID-Spalten aber nicht im Star-Schema?	ok

URL:  
<https://goo.gl/GrwpbN>

# Fakten / Dimensionen (Beispiel)

- Fakten
  - Kosten (Ersatzteilkosten, Mitarbeiterkosten)
- Dimensionen
  - Mitarbeiter (Kostenstelle)
  - Zeit (Anfangstag, Endtag)
  - Reparatur
  - Kunde
  - Fahrzeug

# Klassifikationsschema (Beispiel)



Hierarchie ist immer N:1 Beziehung.  
1:1 entspricht überflüssiger Hierarchiestufe.

# N:M Beziehung

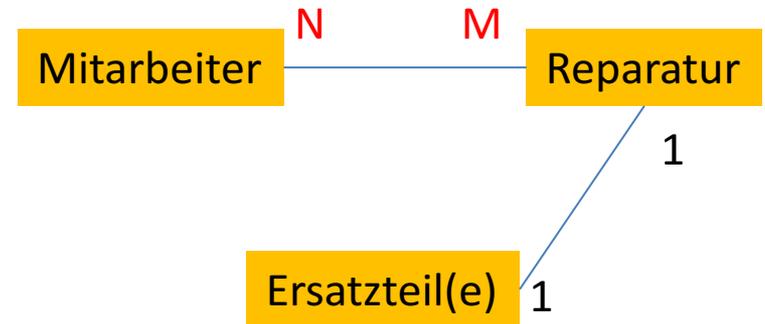
**TO:**  
NAME:  
COMPANY NAME:  
ADDRESS:  
PHONE:

**SHIP TO:**  
NAME:  
COMPANY NAME:  
ADDRESS:  
PHONE:

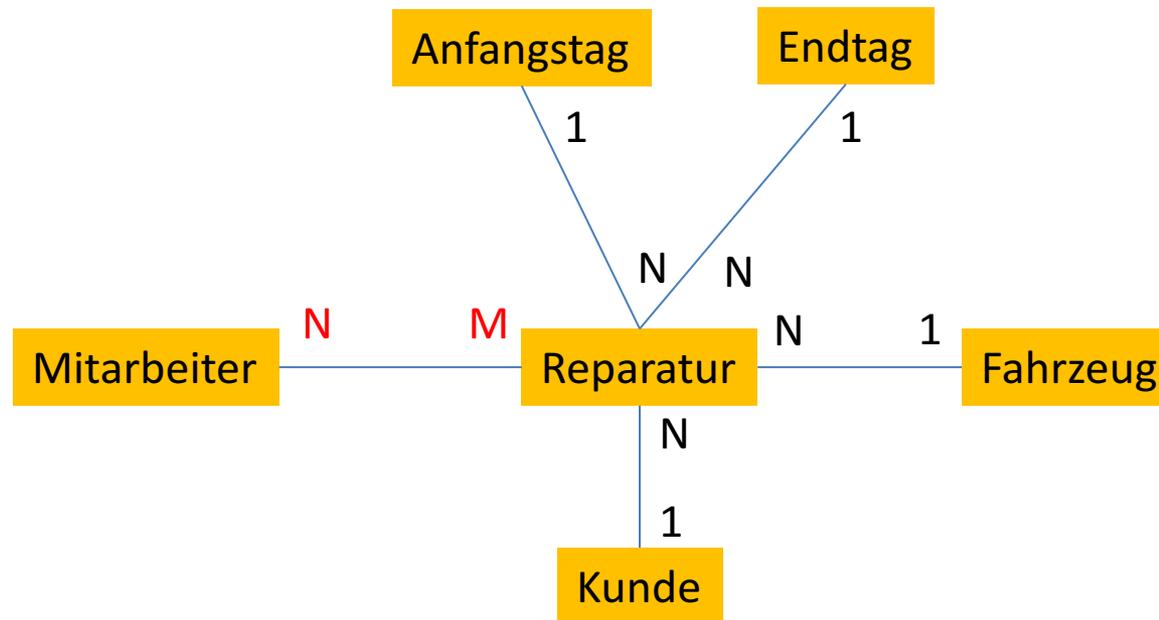
COMMENTS / SPECIAL INSTRUCTIONS:

QUANTITY	DESCRIPTION	UNIT PRICE	TOTAL
	Arbeitszeit, Meier		1000
	Arbeitszeit, Müller		700
	Ersatzteile		200
SUBTOTAL:			
SALES TAX:			
SHIPPING & HANDLING:			1900
TOTAL DUE:			

AUTHORIZED SIGNATURE \_\_\_\_\_

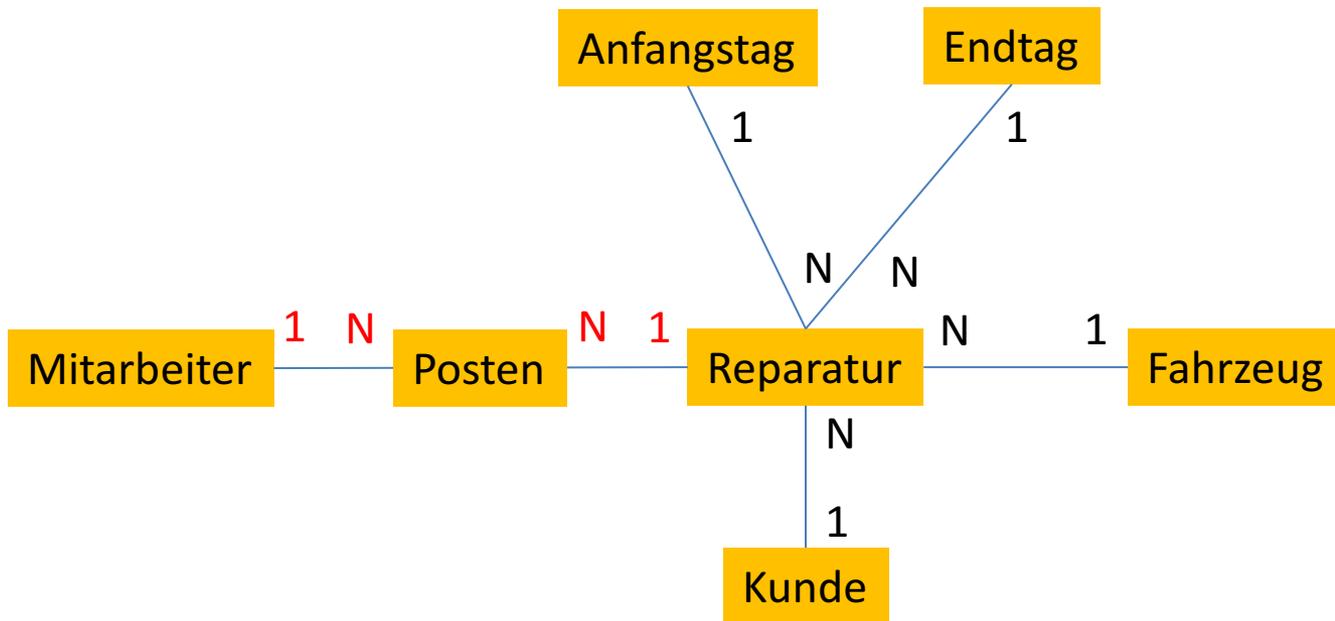


# Kardinalitäten

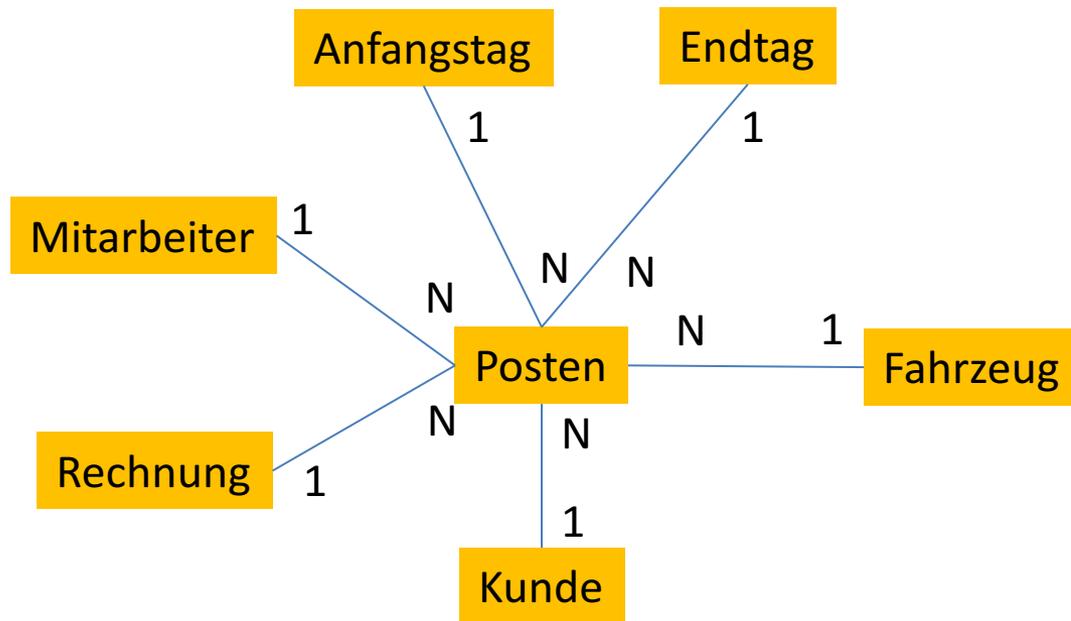


Schlecht: Das erzeugt eine **N:M Beziehung** zwischen der Faktentabelle und Dimensionstabelle. Beeinflusst die Query-Lesbarkeit, Performance, und Struktur negativ.

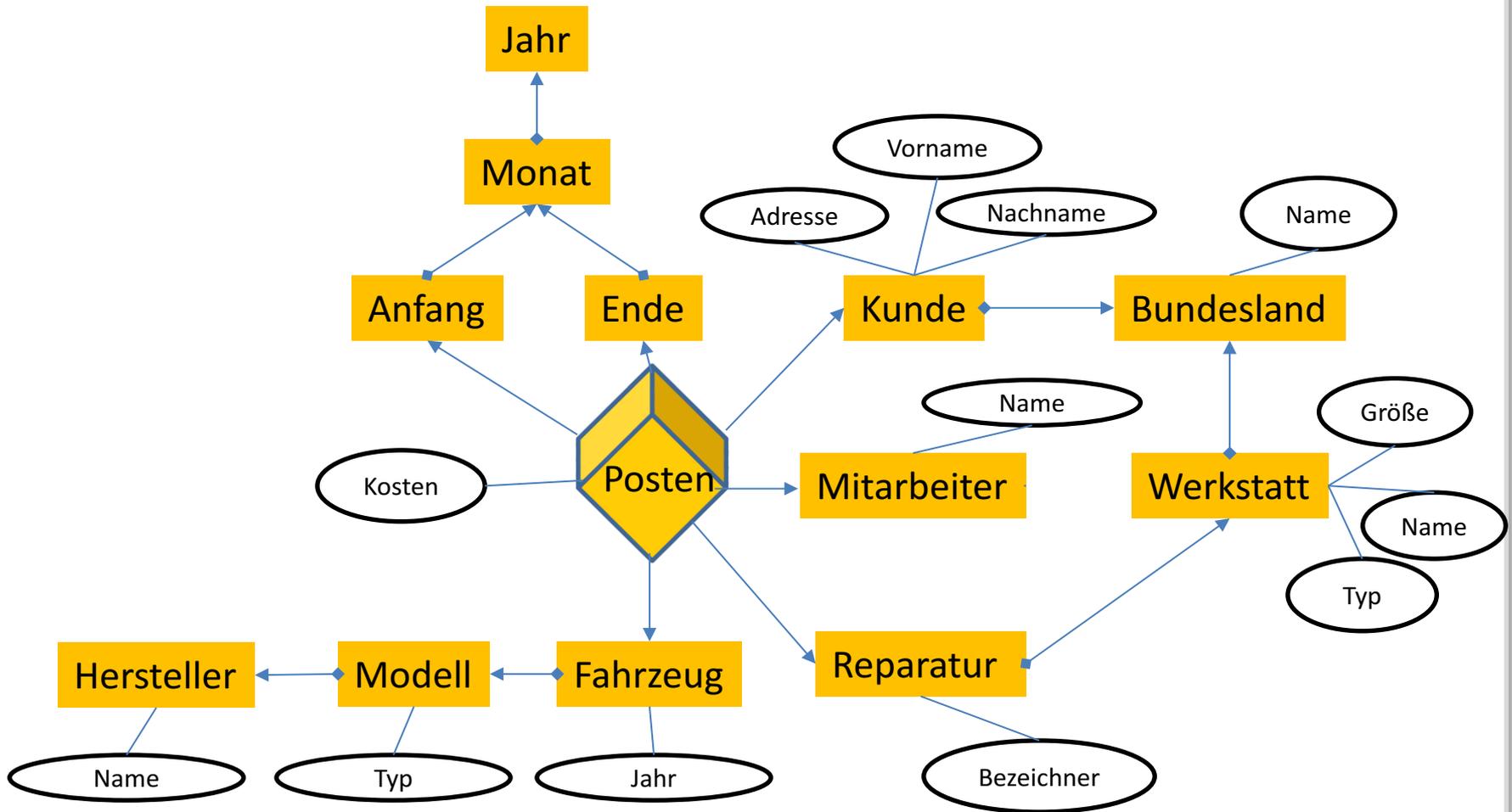
# Transformation (Falsch)



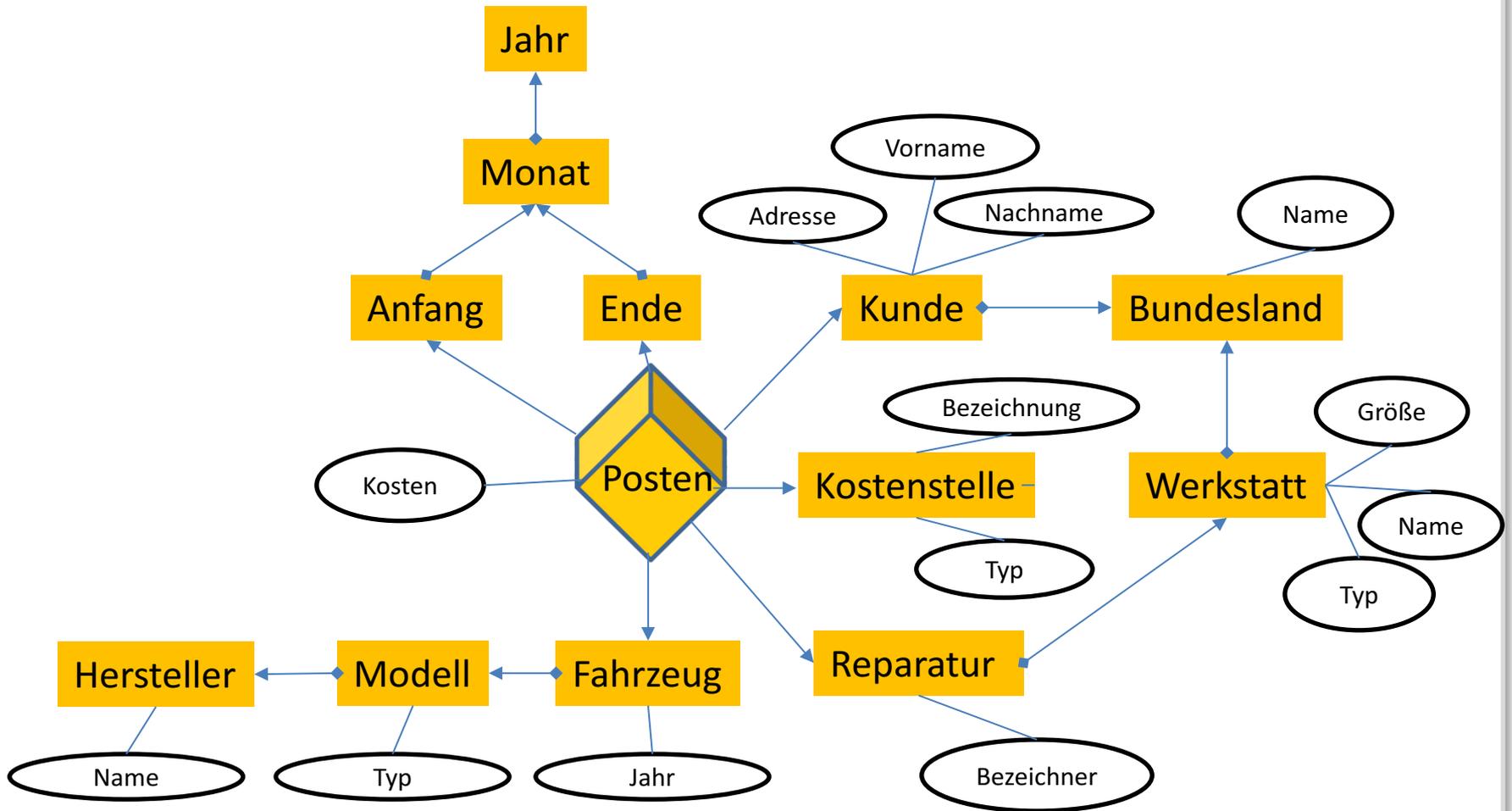
# Transformation (Richtig)



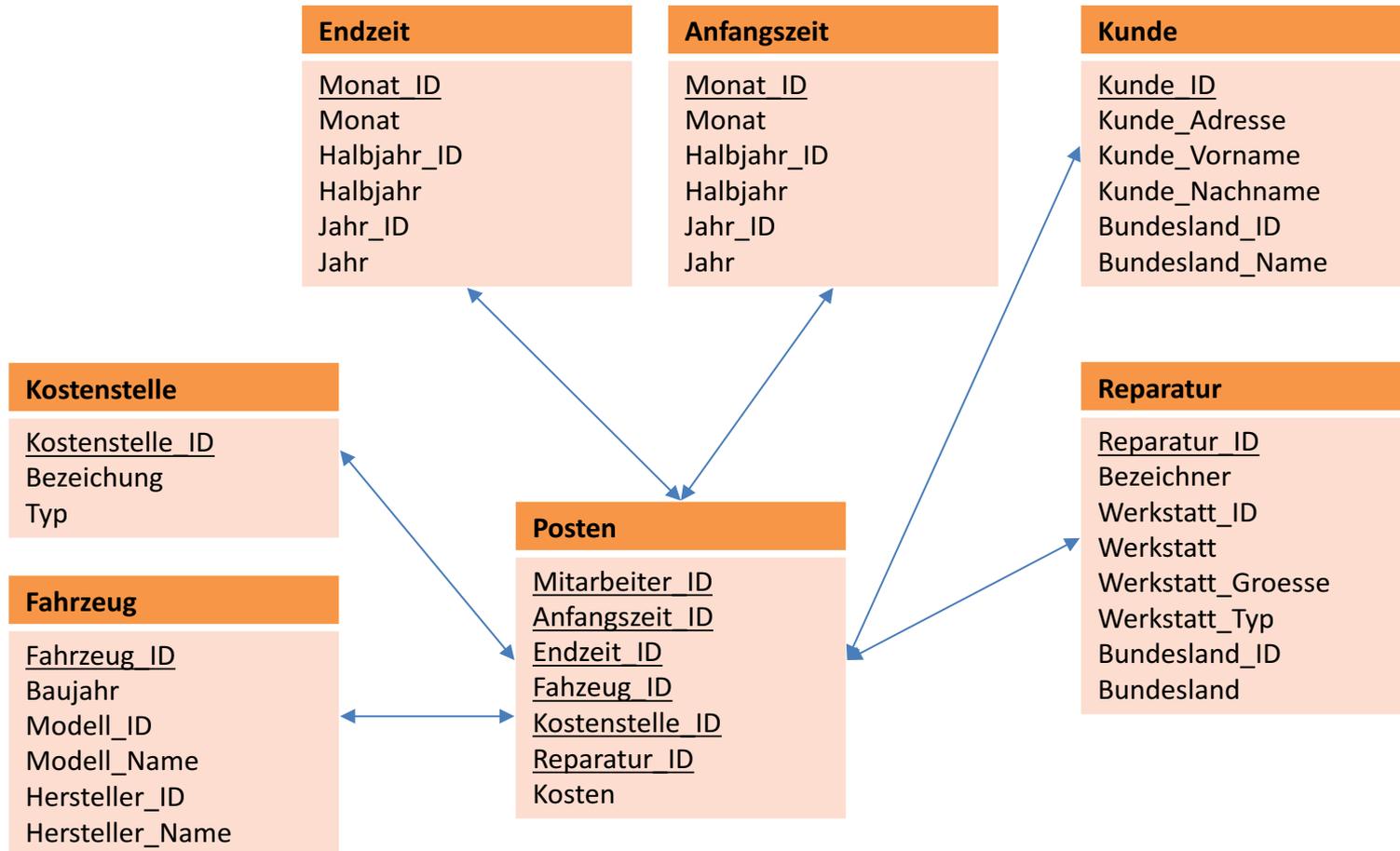
# M/ER Diagramm



# Besser: Kostenstelle



# Star Schema



# 1. Anfrage (Star-Schema)

```
1. SELECT sum(p.kosten)
   FROM Posten p JOIN Kostenstelle k ON ...
   GROUP BY p.Reparatur_id
   HAVING SUM(k.typ) > 1

// Typ speichert zB 1 für Mitarbeiter und
// 0 für Ersatzteil
```

## 2. Anfrage (Star-Schema)

```
SELECT r.bundesland_id,  
       f.hersteller_id,  
       sum(p.kosten)  
FROM   Posten p JOIN Fahrzeug f ON ...  
       JOIN Reparatur r ON ...  
       JOIN Kostenstelle k ON ...  
WHERE  k.typ = 0 // Ersatzteile  
GROUP BY CUBE(r.bundesland_id, f.hersteller_id)  
ORDER BY r.bundesland_id, f.hersteller_id
```

## 3. Anfrage (Star-Schema)

```
SELECT r.bundesland_id, z.jahr_id, z.monat_id,
       sum(p.kosten)
FROM   Posten p JOIN Endzeit z ON ...
       JOIN Reparatur r ON ...
       JOIN Kostenstelle k ON ...
WHERE  k.typ = 1 // Mitarbeiter
GROUP BY GROUPING SET(
         (r.bundesland_id, z.jahr_id, z.monat_id),
         (r.bundesland_id, z.jahr_id),
         ())
ORDER BY z.jahr_id, z.monat_id
```

Fragen?