TECHNISCHE UNIVERSITÄT
KAISERSLAUTERN

Schemamerging und -mapping

Seminar Informationsqualität und
-integration Stefan Hühner, 30.06.2006

Überblick

Schemaintegration / Einleitung

Ziele

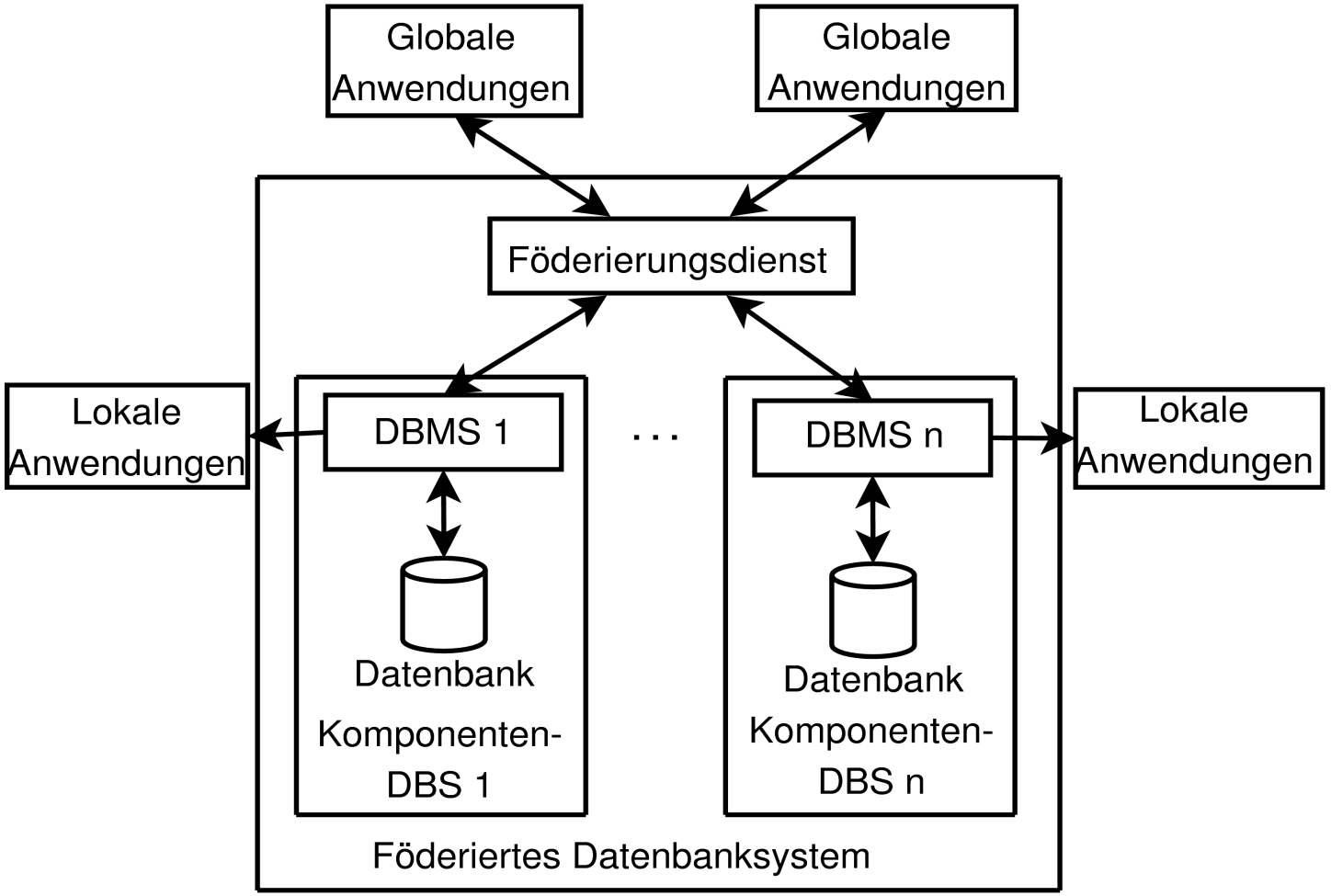
Integrationskonflikte

Integrationstechniken

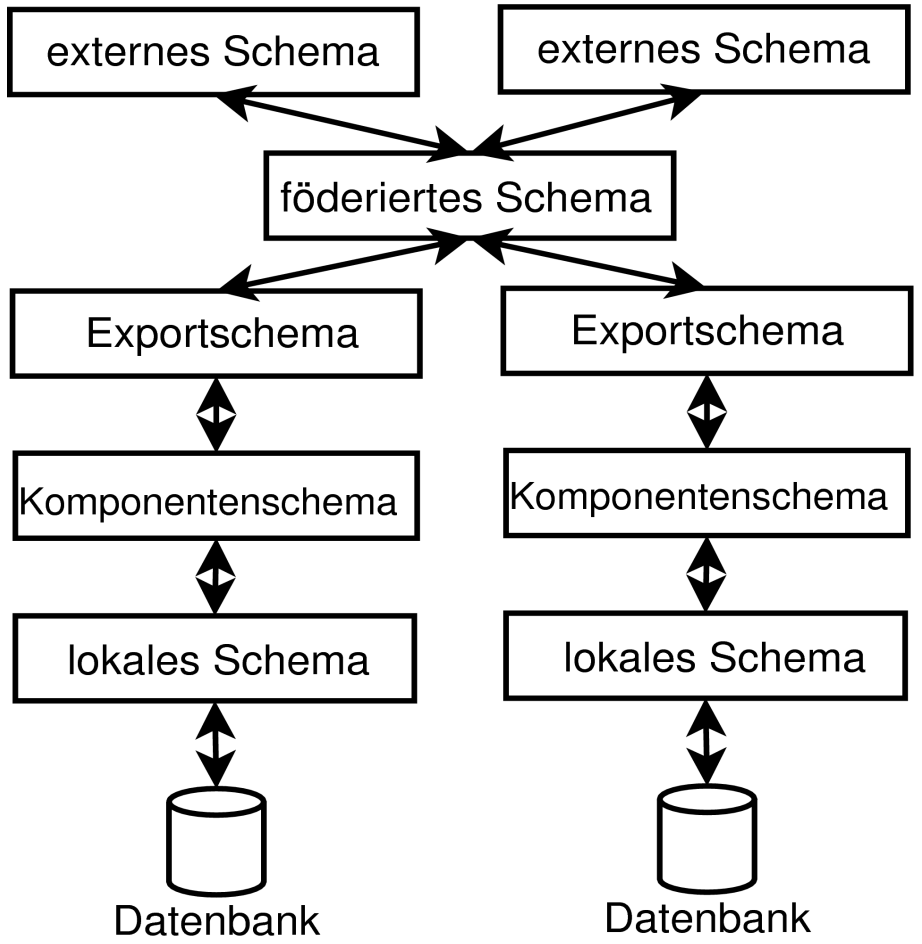
Multidatenbankanfragesprachen

Anfragebearbeitung / -optimierung

Schemaintegration - Umfeld



Schemaintegration



Schemaintegration - Ziele

Vollständigkeit & Korrektheit

Minimalität

Verständlichkeit

Integrationskonflikte

- Heterogenitätskonflikte
  - unterschiedliche Datenmodelle
- Strukturelle Konflikte
  - z.B. verschiedene Normalformen
- Extensionale Konflikte
- Beschreibungskonflikte
  - Name, Vorname <-> Vorname Name
- Datenkonflikte

Integrationstechniken

Zusicherungsbasierte Integration

- Zusammenhänge analysieren

- Element-Korrespondenzen
- Element-Attribut-Korrespondenzen
- Pfad-Korrespondenzen

- Integrationsregeln

Integrationstechniken - Beispiel

1.) Artikel_mech ∩ Artikel_elektrisch
1.a) Sachnr. = Sachnr.
1.b) Benennung = Benennung

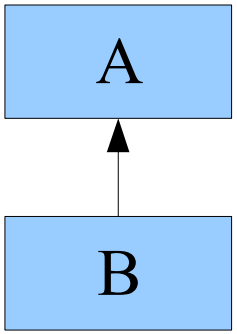
Artikel_mech		Artikel_elektrisch		Artikel_fusioniert
• Sachnr.		• Sachnr.		• Sachnr.
• Benennung		• Benennung		• Benennung
• M-spez. Attr. 1	+	• E-spez. Attr. 1	=	• E-spez. Attr. 1
• M-spez. Attr. 2		• E-spez. Attr. 2		• E-spez. Attr. 2
				• M-spez. Attr. 1
				• M-spez. Attr. 2

Upward Inheritance

Integration von Klassenhierarchien

$$A = B$$

$$A \supseteq B$$



Multidatenbank-Anfragesprachen

mehrere Datenbanken pro Abfrage

Restrukturierung

Metadaten wie Daten behandeln

Abwärtskompatibel mit SQL

dynamische Anfrageschemata

Multidatenbank-Anfragesprachen

- SQL-Sichten
- AJAX
- SchemaSQL
- MQL
- FISQL

SQL-Sichten

o Restrukturierung möglich
++ in jedem DBMS vorhanden

- nur eine DB als Quelle
- statische Struktur
- keine dynamischen Sichten

AJAX

deklarative, erweiterbare Sprache
Fokus ist Datentransformation

Ablauf: Kette von Transformationen
gerichteter azyklischer Graph

Basistransformationen:

- Mapping
- SQL-join / -union

- Matching
- Clustering
- Merging

AJAX - Matching

kartesisches Produkt + ext. Funktionen

Ziel: finden “passender” Tupel

- nicht immer “exact match” möglich
- “fuzzy match” / Abstandsfunktion

Problempotential

SchemaSQL

Variablen in SQL:

- rel — Tupel einer Relation

In SchemaSQL neu:

- db:rel — Tupel einer Relation
- → — Datenbanknamen
- db → — Relationsnamen
- db:rel → — Attributnamen
- db:rel.attr — Werte eines Attributes

SchemaSQL - Beispiel1

db2

zeit

projekt	heinz	hugo
P1	1,5	2,0
P2	2,5	3,0

db1

zeit

projekt	mitarbeiter	arbeitsstunden
P1	heinz	1,5
P1	hugo	2,0
P2	heinz	2,5
P2	hugo	3,0

```
SELECT Z.projekt, S, Z.S
FROM db2:Zeit Z, db2:Zeit → S
WHERE S <> 'projekt'
```

SchemaSQL - Beispiel2

db2

zeit

projekt	heinz	hugo
P1	1,5	2,0
P2	2,5	3,0

db1

zeit

projekt	mitarbeiter	arbeitsstunden
P1	heinz	1,5
P1	hugo	2,0
P2	heinz	2,5
P2	hugo	3,0

```
CREATE VIEW 1zu2::zeit(projekt, S) AS
SELECT Z1.projekt, Z1.arbeitsstunden
FROM db1:zeit Z1, Z1.mitarbeiter S
```

SchemaSQL

+ flexible Restrukturierung
+ Integration von Metadaten
+ dynamische Schemata
+ “Transposition”

- nur eine Spalte als Metadaten
- Umdefinition des “view” Kontrukts
- keine geschachtelten Sichten

FISQL

Algebra FIRA als Basis
Erweiterung von SQL/RA

+ geschachtelte Anfragen
+ Allgemeingültigkeit
+ Abwärtskompatibilität

FISQL - Beispiel

db2

zeit

projekt	heinz	hugo
P1	1,5	2,0
P2	2,5	3,0

db1

zeit

projekt	mitarbeiter	arbeitsstunden
P1	heinz	1,5
P1	hugo	2,0
P2	heinz	2,5
P2	hugo	3,0

```
SELECT Z.projekt as 'projekt',
Z.arbeitsstunden ON Z.mitarbeiter
INTO '1zu2'
FROM db1:zeit as Z
```

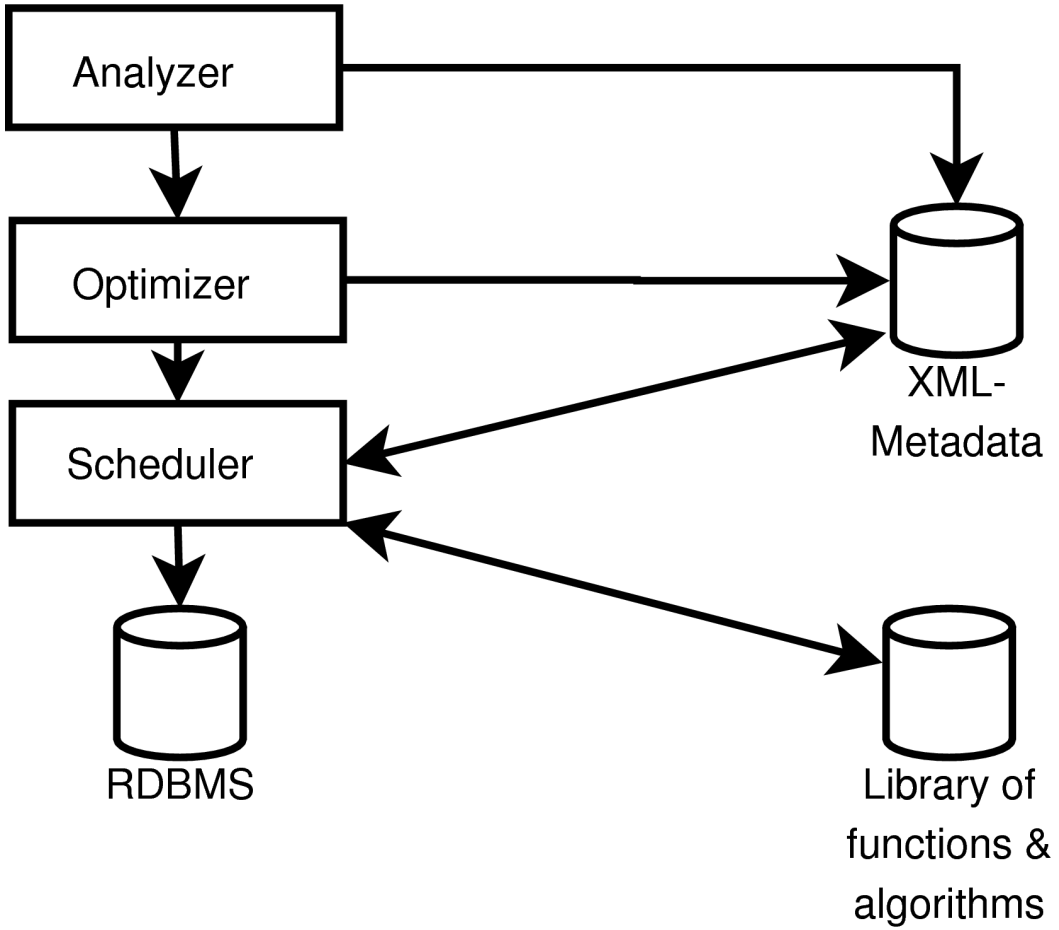
Anfragebearbeitung

Ablauf einer Anfrage

Unterschiede zu SQL

Optimierungsmöglichkeiten

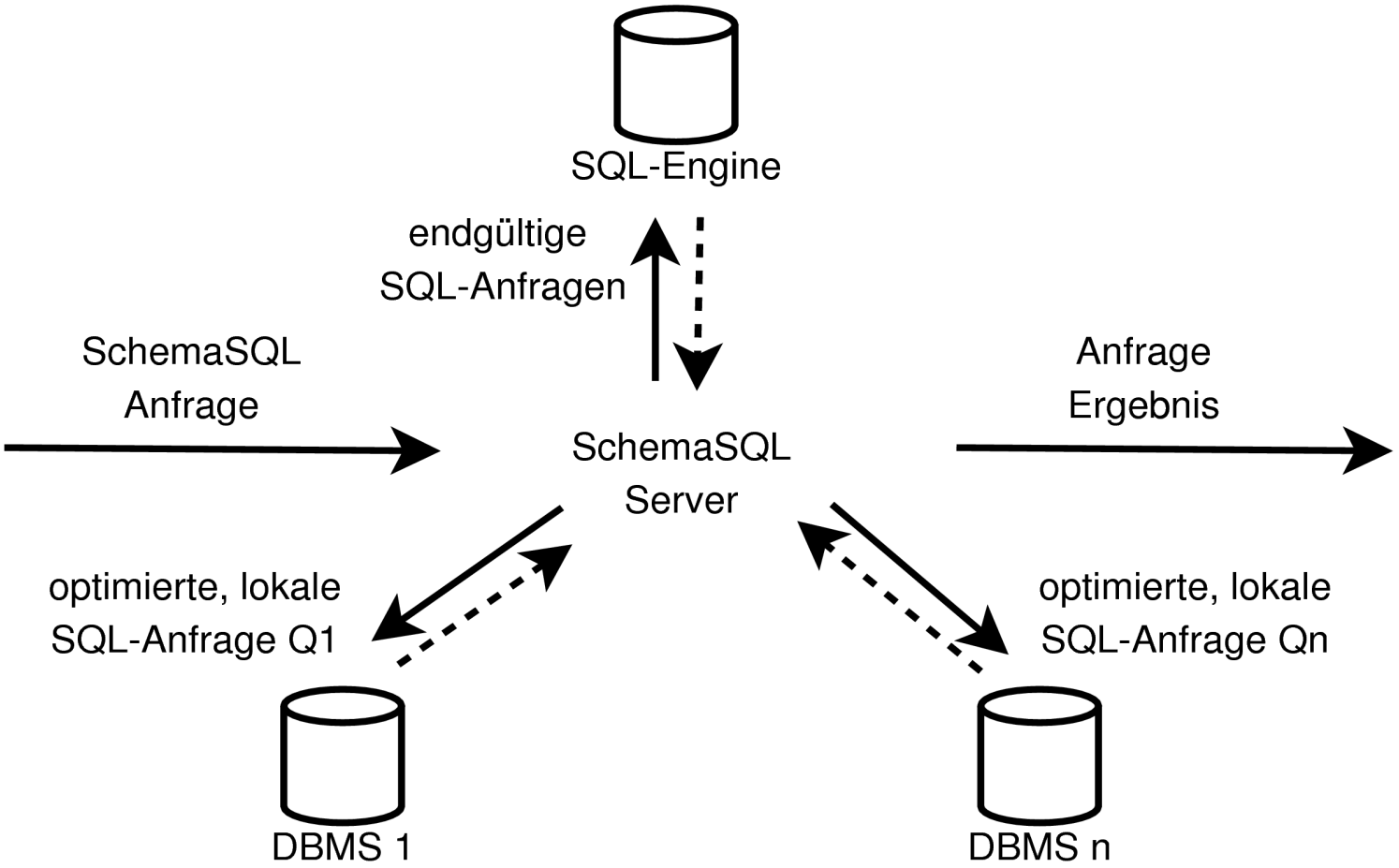
AJAX - Architektur



Anfrageoptimierung - AJAX

- Funktions-Umordnung
- Short-Circuited Computation
- Cached Computation
- Parallele Auswertung

Anfragebearbeitung - SchemaSQL



Anfrageoptimierung - SchemaSQL

- Umordnung von Selektionen
- Batch-Abfragen
- Parallele Abfragen

- Minimierung / Zusammenfassung von Zwischenergebnissen

Zusammenfassung

Integrationsziele

Integrationskonflikte

Multidatenbank-Anfragesprachen

Anfragebearbeitung / -optimierung

Fragen ?