

intro

mcluster
bdc
problem

related

mdc
others
conclusion

ideas

concept
precondition
abstraction
transformation
cycles
full auto
example
extensions

tests

tpch
benchmarks

Automatisierter Entwurf eines Schemas für Bitwise Dimensional Clustering

–Verteidigung–

Steve Göring

16. Februar 2012
V0.9

Steve Göring

intro

mcluster
bdc
problem

related

mdc
others
conclusion

ideas

concept
precondition
abstraction
transformation
cycles
full auto
example
extensions

tests

tpch
benchmarks

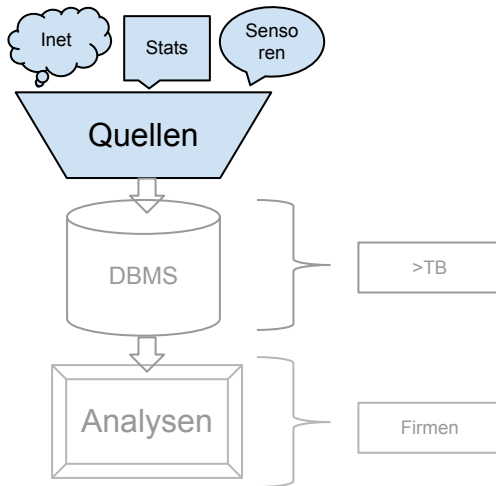
1 intro

2 related

3 ideas

4 tests

Einleitung (0)



intro

- mcluster
- bdc
- problem

related

- mdc
- others
- conclusion

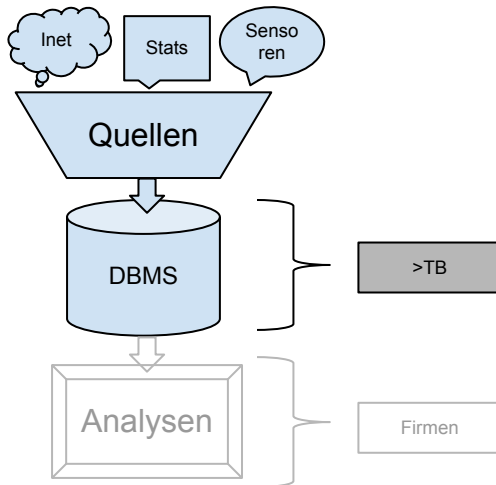
ideas

- concept
- precondition
- abstraction
- transformation
- cycles
- full auto
- example
- extensions

tests

- tpch
- benchmarks

Einleitung (1)



intro

- mcluster
- bdc
- problem

related

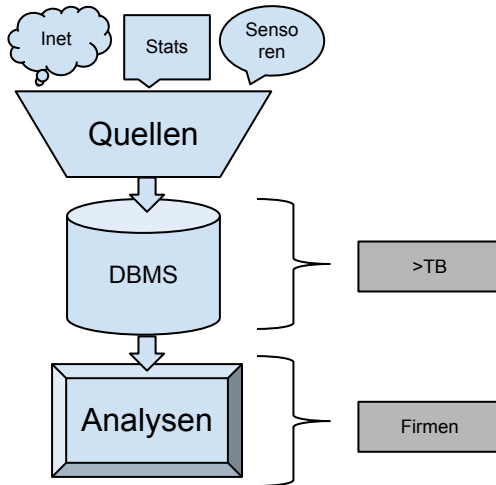
- mdc
- others
- conclusion

ideas

- concept
- precondition
- abstraction
- transformation
- cycles
- full auto
- example
- extensions

tests

- tpch
- benchmarks



intro

- mcluster
- bdc
- problem

related

- mdc
- others
- conclusion

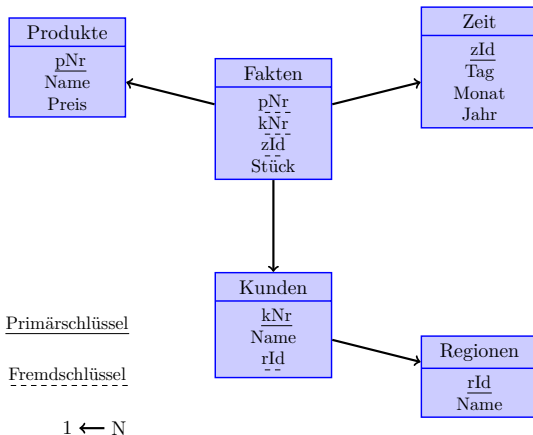
ideas

- concept
- precondition
- abstraction
- transformation
- cycles
- full auto
- example
- extensions

tests

- tpch
- benchmarks

Einleitung - (Online Versand)



$|Kunden| = |Fakten| = |Zeit| = 1000$
 $\rightarrow Kunden \times Fakten \times Zeit = 1 \text{ Mio.}$

Multidimensionales Clustern

intro

mcluster

bdc

problem

related

mdc

others

conclusion

ideas

concept

precondition

abstraction

transformation

cycles

full auto

example

extensions

tests

tpch

benchmarks

Daten eingeteilt in:

- ▷ Dimensionen:
beschreibende Daten
- ▷ Fakten:
Kern/ Ergebniszahlen

Multidimensionales Clustern

intro

mcluster

bdc

problem

related

mdc

others

conclusion

ideas

concept

precondition

abstraction

transformation

cycles

full auto

example

extensions

tests

tpch

benchmarks

Daten eingeteilt in:

- ▷ Dimensionen:
beschreibende Daten
- ▷ Fakten:
Kern/ Ergebniszahlen
 - ◇ geclustert durch Dimensionen
 - ◇ Daten mit ähnlichen Dims dicht beieinander
 - ◇ Clusterinfos zur Datenreduktion

Bitwise Dimensional Clustering

intro

mcluster

bdc

problem

related

mdc

others

conclusion

ideas

concept

precondition

abstraction

transformation

cycles

full auto

example

extensions

tests

tpch

benchmarks

Bitwise Dimensional Clustering (BDC):

- ▷ multidimensionales Speichersystem für Columnstores (Vektorwise)
- ▷ Clusterungsproblem → Ordnungsproblem

Grundidee von BDC:

- ▷ Dimensionen → Integerzahlen
- ▷ Tabellen geclustert nach Dimensionsauswahl

intro

mcluster

bdc

problem

related

mdc

others

conclusion

ideas

conecpt

precondition

abstraction

transformation

cycles

full auto

example

extensions

tests

tpch

bechmarks

Festlegung von Dimensionen:

intro

mcluster

bdc

problem

related

mdc

others

conclusion

ideas

concept

precondition

abstraction

transformation

cycles

full auto

example

extensions

tests

tpch

benchmarks

Festlegung von Dimensionen:

- ▷ Tabellenname
- ▷ Dimensionsspalte
- ▷ *cbits*

intro

mcluster

bdc

problem

related

mdc

others

conclusion

ideas

concept

precondition

abstraction

transformation

cycles

full auto

example

extensions

tests

tpch

bechmarks

Festlegung von Dimensionen:

- ▷ Tabellenname
- ▷ Dimensionsspalte
- ▷ *cbits*

BDC-geclusterte Tabelle:

intro

mcluster

bdc

problem

related

mdc

others

conclusion

ideas

concept

precondition

abstraction

transformation

cycles

full auto

example

extensions

tests

tpch

benchmarks

Festlegung von Dimensionen:

- ▷ Tabellenname
- ▷ Dimensionsspalte
- ▷ *cbits*

BDC-geclusterte Tabelle:

- ▷ Pfad zur jeweiligen Dimension
- ▷ Bitmaske für jede Dimension
- ▷ *iobits*

Problemstellung

intro

mcluster

bdc

problem

related

mdc

others

conclusion

ideas

concept

precondition

abstraction

transformation

cycles

full auto

example

extensions

tests

tpch

bechmarks

▷ komplexere Schemas → Dimensionsauswahl schwerer

Problemstellung

intro

mcluster

bdc

problem

related

mdc

others

conclusion

ideas

concept

precondition

abstraction

transformation

cycles

full auto

example

extensions

tests

tpch

benchmarks

- ▷ komplexere Schemas → Dimensionsauswahl schwerer
- ▷ viele Faktoren beeinflussen Clusterung

Problemstellung

intro

mcluster

bdc

problem

related

mdc

others

conclusion

ideas

concept

precondition

abstraction

transformation

cycles

full auto

example

extensions

tests

tpch

benchmarks

- ▷ komplexere Schemas → Dimensionsauswahl schwerer
 - ▷ viele Faktoren beeinflussen Clusterung
- ⇒ automatisches Erzeugen des Schemas

intro

mcluster
bdc
problem

related

mdc
others
conclusion

ideas

concept
precondition
abstraction
transformation
cycles
full auto
example
extensions

tests

tpch
benchmarks

MDC in DB2 ¹

- ▷ Lightstone, Bhattacharjee ²
- ▷ automatisiertes Verfahren
- ▷ als Ratgeber implementiert

für Dimensionen Analyse:

- ▷ Daten
- ▷ Anfragen (GROUP BY, ORDER BY, WHERE)

¹“Database Partitioning, Table Partitioning, and MDC for DB2 9“, 2009

²“Automated Design of Multidimensional Clustering Tables for Relational Databases“, 2004

Verwandte Themen - andere Verfahren

intro

mcluster
bdc
problem

related

mdc
others
conclusion

ideas

concept
precondition
abstraction
transformation
cycles
full auto
example
extensions

tests

tpch
bechmarks

- ▷ fast jedes DBMS: Indexassistenten
- ▷ Grundlage: Daten/Anfrageanalyse

Steve Göring

intro

mcluster

bdc

problem

related

mdc

others

conclusion

ideas

conecpt

precondition

abstraction

transformation

cycles

full auto

example

extensions

tests

tpch

bechmarks

▷ Datenanalyse

Steve Göring

intro

mcluster

bdc

problem

related

mdc

others

conclusion

ideas

concept

precondition

abstraction

transformation

cycles

full auto

example

extensions

tests

tpch

bechmarks

- ▷ Datenanalyse :)
- ▷ Anfrageanalyse

Steve Göring

intro

mcluster

bdc

problem

related

mdc

others

conclusion

ideas

concept

precondition

abstraction

transformation

cycles

full auto

example

extensions

tests

tpch

benchmarks

- ▷ Datenanalyse :)
- ▷ Anfrageanalyse :(

Steve Göring

intro

mcluster
bdc
problem

related

mdc
others
conclusion

ideas

concept
precondition
abstraction
transformation
cycles
full auto
example
extensions

tests

tpch
bechmarks

- ▷ Datenanalyse :)
- ▷ Anfrageanalyse :(

Ziel:

- ▷ innerhalb von Vectorwise
- ▷ auf BDC abgestimmt

Steve Göring

intro

mcluster
bdc
problem

related

mdc
others
conclusion

ideas

concept
precondition
abstraction
transformation
cycles
full auto
example
extensions

tests

tpch
benchmarks

▷ Dimensionstabellen:

intro

mcluster

bdc

problem

related

mdc

others

conclusion

ideas

concept

precondition

abstraction

transformation

cycles

full auto

example

extensions

tests

tpch

benchmarks

- ▷ Dimensionstabellen:
 - ◇ Tabellen, die keine Fremdschlüsselbeziehungen zu anderen besitzen

intro

mcluster
bdc
problem

related

mdc
others
conclusion

ideas

concept
precondition
abstraction
transformation
cycles
full auto
example
extensions

tests

tpch
bechmarks

- ▷ Dimensionstabellen:
 - ◇ Tabellen, die keine Fremdschlüsselbeziehungen zu anderen besitzen
 - ◇ benutzen zur Clusterung eigene Dimension

intro

mcluster
bdc
problem

related

mdc
others
conclusion

ideas

concept
precondition
abstraction
transformation
cycles
full auto
example
extensions

tests

tpch
bechmarks

- ▷ Dimensionstabellen:
 - ◇ Tabellen, die keine Fremdschlüsselbeziehungen zu anderen besitzen
 - ◇ benutzen zur Clusterung eigene Dimension
- ▷ andere Tabellen:

intro

mcluster
bdc
problem

related

mdc
others
conclusion

ideas

concept
precondition
abstraction
transformation
cycles
full auto
example
extensions

tests

tpch
bechmarks

- ▷ Dimensionstabellen:
 - ◇ Tabellen, die keine Fremdschlüsselbeziehungen zu anderen besitzen
 - ◇ benutzen zur Clusterung eigene Dimension
- ▷ andere Tabellen:
 - ◇ benutzen Informationen der Nachbartabellen zum clustern

intro

mcluster
bdc
problem

related

mdc
others
conclusion

ideas

concept
precondition
abstraction
transformation
cycles
full auto
example
extensions

tests

tpch
benchmarks

- ▷ Dimensionstabellen:
 - ◇ Tabellen, die keine Fremdschlüsselbeziehungen zu anderen besitzen
 - ◇ benutzen zur Clusterung eigene Dimension
- ▷ andere Tabellen:
 - ◇ benutzen Informationen der Nachbartabellen zum clustern
- ▷ Abarbeitung von “Außen nach Innen“

intro

mcluster
bdc
problem

related

mdc
others
conclusion

ideas

concept
precondition
abstraction
transformation
cycles
full auto
example
extensions

tests

tpch
benchmarks

- ▷ Dimensionstabellen:
 - ◇ Tabellen, die keine Fremdschlüsselbeziehungen zu anderen besitzen
 - ◇ benutzen zur Clusterung eigene Dimension
- ▷ andere Tabellen:
 - ◇ benutzen Informationen der Nachbartabellen zum clustern
- ▷ Abarbeitung von “Außen nach Innen“
- ▷ eventuell vorher Transformation nötig

intro

mcluster

bdc

problem

related

mdc

others

conclusion

ideas

concept

precondition

abstraction

transformation

cycles

full auto

example

extensions

tests

tpch

benchmarks

▷ Informationen über Schema:

intro

mcluster
bdc
problem

related

mdc
others
conclusion

ideas

concept
precondition
abstraction
transformation
cycles
full auto
example
extensions

tests

tpch
benchmarks

- ▷ Informationen über Schema:
 - ◇ Tabellennamen

intro

mcluster
bdc
problem

related

mdc
others
conclusion

ideas

concept
precondition
abstraction
transformation
cycles
full auto
example
extensions

tests

tpch
bechmarks

- ▷ Informationen über Schema:
 - ◇ Tabellennamen
 - ◇ Fremdschlüssel / Primärschlüssel Beziehungen

intro

mcluster
bdc
problem

related

mdc
others
conclusion

ideas

concept
precondition
abstraction
transformation
cycles
full auto
example
extensions

tests

tpch
bechmarks

- ▷ Informationen über Schema:
 - ◇ Tabellennamen
 - ◇ Fremdschlüssel / Primärschlüssel Beziehungen
- ▷ Statistische Informationen: (Histogramm/ Anfragen)

intro

mcluster
bdc
problem

related

mdc
others
conclusion

ideas

concept
precondition
abstraction
transformation
cycles
full auto
example
extensions

tests

tpch
bechmarks

- ▷ Informationen über Schema:
 - ◇ Tabellennamen
 - ◇ Fremdschlüssel / Primärschlüssel Beziehungen
- ▷ Statistische Informationen: (Histogramm/ Anfragen)
 - ◇ Min/Max/Count

intro

mcluster
bdc
problem

related

mdc
others
conclusion

ideas

concept
precondition
abstraction
transformation
cycles
full auto
example
extensions

tests

tpch
bechmarks

- ▷ Informationen über Schema:
 - ◇ Tabellennamen
 - ◇ Fremdschlüssel / Primärschlüssel Beziehungen
- ▷ Statistische Informationen: (Histogramm/ Anfragen)
 - ◇ Min/Max/Count
 - ◇ Blockgröße

Schema \rightarrow Graph

▷ Schema

intro

mcluster

bdc

problem

related

mdc

others

conclusion

ideas

conecpt

precondition

abstraction

transformation

cycles

full auto

example

extensions

tests

tpch

bechmarks

Schema \rightarrow Graph

▷ Schema \rightarrow

intro

mcluster

bdc

problem

related

mdc

others

conclusion

ideas

concept

precondition

abstraction

transformation

cycles

full auto

example

extensions

tests

tpch

benchmarks

Schema \rightarrow Graph

intro

mcluster

bdc

problem

related

mdc

others

conclusion

ideas

concept

precondition

abstraction

transformation

cycles

full auto

example

extensions

tests

tpch

bechmarks

- ▷ Schema \rightarrow erweiterter gerichteter Multigraph
 $G = (V, E, \Delta)$, mit:

Schema \rightarrow Graph

intro

mcluster

bdc

problem

related

mdc

others

conclusion

ideas

concept

precondition

abstraction

transformation

cycles

full auto

example

extensions

tests

tpch

benchmarks

▷ Schema \rightarrow erweiterter gerichteter Multigraph

$G = (V, E, \Delta)$, mit:

◇ $V = \{ \textit{Tabellennamen} \}$

Schema \rightarrow Graph

intro

mcluster
bdc
problem

related

mdc
others
conclusion

ideas

concept
precondition

abstraction

transformation
cycles
full auto
example
extensions

tests

tpch
bechmarks

▷ Schema \rightarrow erweiterter gerichteter Multigraph

$G = (V, E, \Delta)$, mit:

◇ $V = \{ \text{Tabellennamen} \}$

◇ $E = \{ (x, y) \mid x \text{ hat FK zum PK von } y \}$!Multimenge!

Schema \rightarrow Graph

intro

mcluster
bdc
problem

related

mdc
others
conclusion

ideas

concept
precondition

abstraction

transformation
cycles
full auto
example
extensions

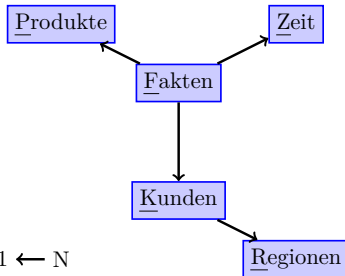
tests

tpch
benchmarks

▷ Schema \rightarrow erweiterter gerichteter Multigraph

$G = (V, E, \Delta)$, mit:

- ◇ $V = \{ \text{Tabellennamen} \}$
- ◇ $E = \{ (x, y) \mid x \text{ hat FK zum PK von } y \}$!Multimenge!
- ◇ Δ - Funktion: Pfadspeicherungsfunktion
 $\Delta : E \mapsto \text{Pot}(\text{Pfade})$



intro

mcluster

bdc

problem

related

mdc

others

conclusion

ideas

concept

precondition

abstraction

transformation

cycles

full auto

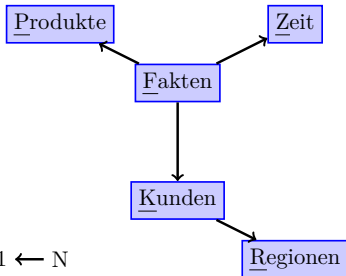
example

extensions

tests

tpch

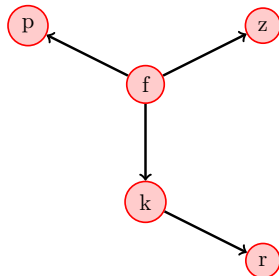
benchmarks

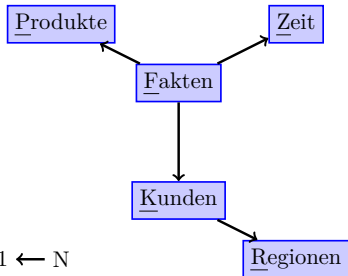


1 ← N

▷ $V = \{p, f, z, k, r\}$

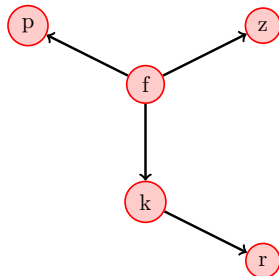
Beispiel

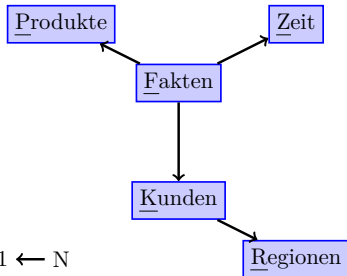



 $1 \leftarrow N$

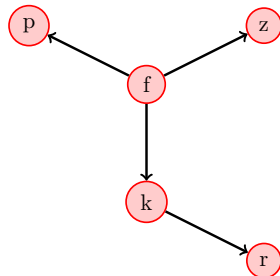
- ▷ $V = \{p, f, z, k, r\}$
- ▷ $E = \{(f, p), (f, z), (f, k), (k, r)\}$

Beispiel




 $1 \leftarrow N$

Beispiel



- ▷ $V = \{p, f, z, k, r\}$
- ▷ $E = \{(f, p), (f, z), (f, k), (k, r)\}$
- ▷ $\Delta(x) = x, \forall x \in E$

Transformationskriterien

intro

mcluster

bdc

problem

related

mdc

others

conclusion

ideas

concept

precondition

abstraction

transformation

cycles

full auto

example

extensions

tests

tpch

benchmarks

Transformationskriterien

intro

mcluster
bdc
problem

related

mdc
others
conclusion

ideas

conecpt
precondition
abstraction

transformation

cycles
full auto
example
extensions

tests

tpch
bechmarks

▷ zusammenhängend

Transformationskriterien

intro

mcluster
bdc
problem

related

mdc
others
conclusion

ideas

conecpt
precondition
abstraction

transformation

cycles
full auto
example
extensions

tests

tpch
bechmarks

▷ zusammenhängend → mehrmaliges Anwenden

Transformationskriterien

intro

mcluster
bdc
problem

related

mdc
others
conclusion

ideas

concept
precondition
abstraction

transformation

cycles
full auto
example
extensions

tests

tpch
bechmarks

- ▷ zusammenhängend → mehrmaliges Anwenden
- ▷ Eigenkanten

Transformationskriterien

intro

mcluster
bdc
problem

related

mdc
others
conclusion

ideas

concept
precondition
abstraction

transformation

cycles
full auto
example
extensions

tests

tpch
bechmarks

- ▷ zusammenhängend → mehrmaliges Anwenden
- ▷ Eigenkanten → entfallen

Transformationskriterien

intro

mcluster
bdc
problem

related

mdc
others
conclusion

ideas

concept
precondition
abstraction

transformation

cycles
full auto
example
extensions

tests

tpch
bechmarks

- ▷ zusammenhängend → mehrmaliges Anwenden
- ▷ Eigenkanten → entfallen
- ▷ Kreise

Transformationskriterien

intro

mcluster
bdc
problem

related

mdc
others
conclusion

ideas

concept
precondition
abstraction
transformation

cycles
full auto
example
extensions

tests

tpch
bechmarks

- ▷ zusammenhängend → mehrmaliges Anwenden
- ▷ Eigenkanten → entfallen
- ▷ Kreise → transformiert

Transformationskriterien

intro

mcluster
bdc
problem

related

mdc
others
conclusion

ideas

concept
precondition
abstraction

transformation

cycles
full auto
example
extensions

tests

tpch
bechmarks

- ▷ zusammenhängend → mehrmaliges Anwenden
- ▷ Eigenkanten → entfallen
- ▷ Kreise → transformiert

⇒ zyklusfreier, zusammenhängender Multigraph $G = (V, E, \Delta)$

Kreise - (allgemein)

intro

mcluster
bdc
problem

related

mdc
others
conclusion

ideas

conecpt
precondition
abstraction
transformation

cycles

full auto
example
extensions

tests

tpch
bechmarks

▷ Tiefensuche (DFS) um Kreise zu finden

Kreise - (allgemein)

intro

mcluster
bdc
problem

related

mdc
others
conclusion

ideas

conecpt
precondition
abstraction
transformation

cycles

full auto
example
extensions

tests

tpch
bechmarks

- ▷ Tiefensuche (DFS) um Kreise zu finden
- ▷ für jeden Knoten eines Kreises:

Kreise - (allgemein)

intro

mcluster
bdc
problem

related

mdc
others
conclusion

ideas

conecpt
precondition
abstraction
transformation

cycles

full auto
example
extensions

tests

tpch
bechmarks

- ▷ Tiefensuche (DFS) um Kreise zu finden
- ▷ für jeden Knoten eines Kreises:
 - ◇ DFS vom Knoten ausführen um neue Pfade zu den Blättern zu finden

Kreise - (allgemein)

intro

mcluster
bdc
problem

related

mdc
others
conclusion

ideas

conecpt
precondition
abstraction
transformation
cycles
full auto
example
extensions

tests

tpch
bechmarks

- ▷ Tiefensuche (DFS) um Kreise zu finden
- ▷ für jeden Knoten eines Kreises:
 - ◇ DFS vom Knoten ausführen um neue Pfade zu den Blättern zu finden
 - ◇ für jeden Pfad: füge Abkürzung in E ein

Kreise - (allgemein)

intro

mcluster
bdc
problem

related

mdc
others
conclusion

ideas

conecpt
precondition
abstraction
transformation
cycles
full auto
example
extensions

tests

tpch
bechmarks

- ▷ Tiefensuche (DFS) um Kreise zu finden
- ▷ für jeden Knoten eines Kreises:
 - ◇ DFS vom Knoten ausführen um neue Pfade zu den Blättern zu finden
 - ◇ für jeden Pfad: füge Abkürzung in E ein
 - ◇ merken des Pfades via Δ -Funktion

Kreise - (allgemein)

intro

mcluster
bdc
problem

related

mdc
others
conclusion

ideas

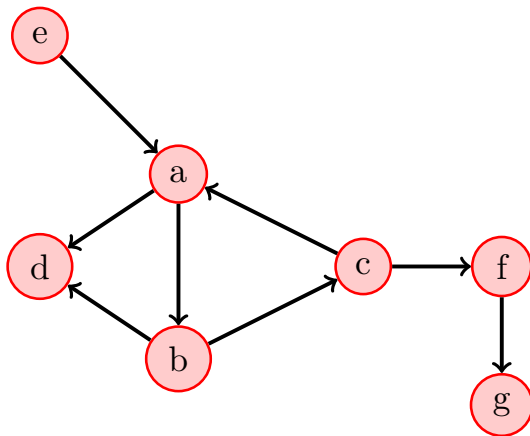
conecpt
precondition
abstraction
transformation
cycles
full auto
example
extensions

tests

tpch
bechmarks

- ▷ Tiefensuche (DFS) um Kreise zu finden
- ▷ für jeden Knoten eines Kreises:
 - ◇ DFS vom Knoten ausführen um neue Pfade zu den Blättern zu finden
 - ◇ für jeden Pfad: füge Abkürzung in E ein
 - ◇ merken des Pfades via Δ -Funktion
- ▷ Entfernen der Kreiskanten

Kreise - (0)



intro

mcluster

bdc

problem

related

mdc

others

conclusion

ideas

concept

precondition

abstraction

transformation

cycles

full auto

example

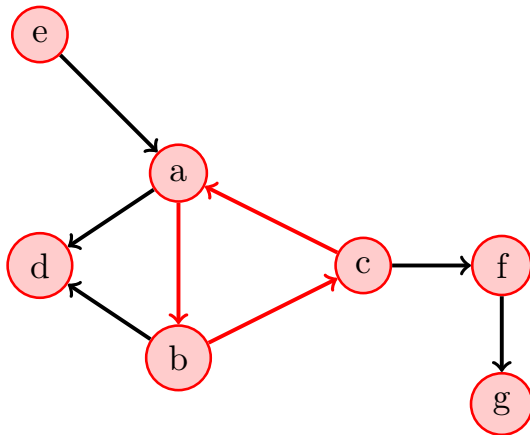
extensions

tests

tpch

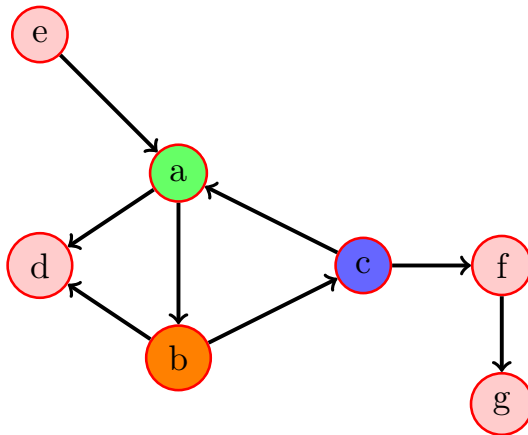
benchmarks

Kreise - (1)



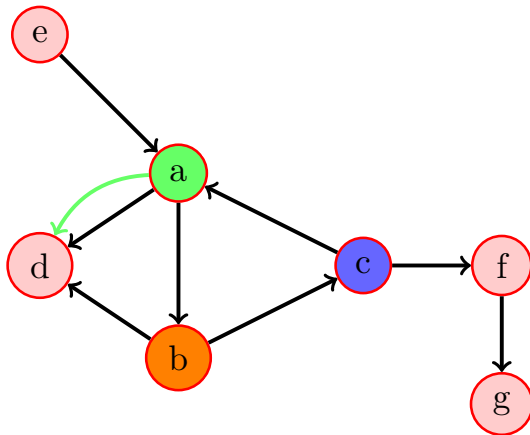
Kreiskanten

Kreise - (2)



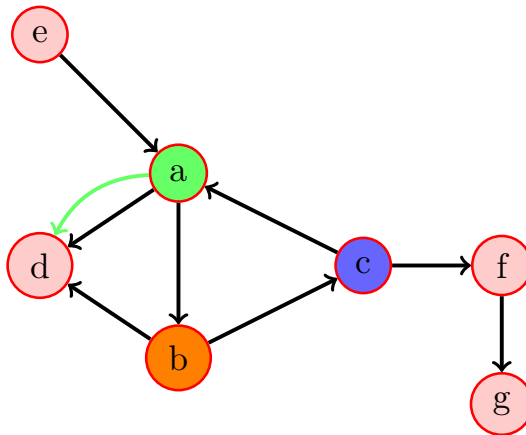
Tiefensuche: Knoten a

Kreise - (2)



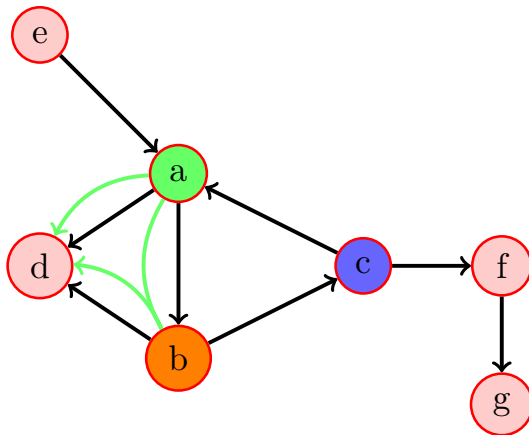
Tiefensuche: Knoten a : Pfad: a, d

Kreise - (2)



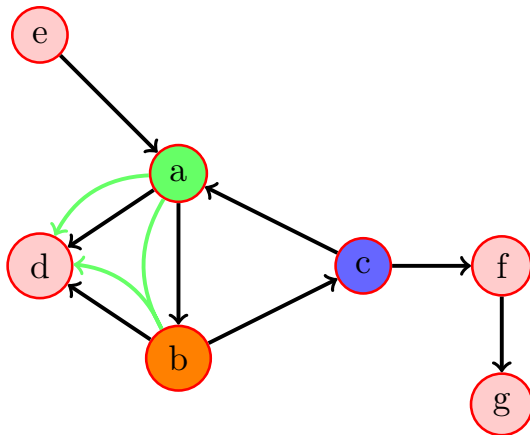
Tiefensuche: Knoten a : Pfad: $a, d \rightarrow$ originale Kante

Kreise - (2)



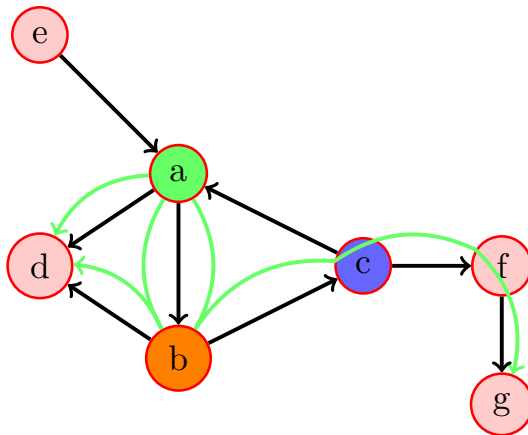
Tiefensuche: Knoten a : Pfad: a, b, d

Kreise - (2)



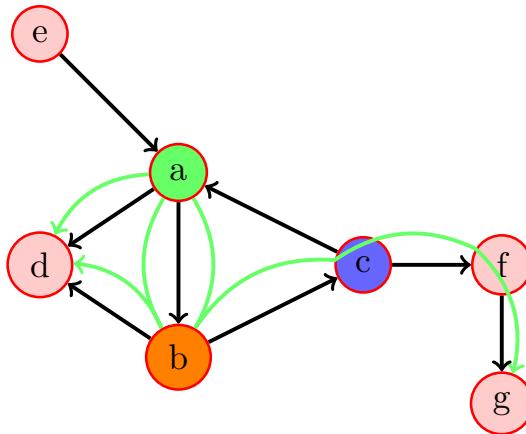
Tiefensuche: Knoten a : Pfad: $a, b, d \rightarrow$ neu (a, d)

Kreise - (2)



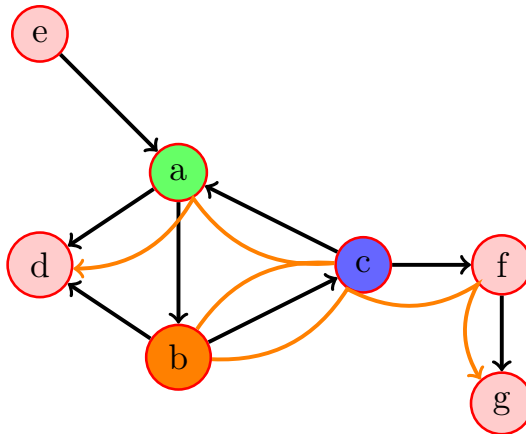
Tiefensuche: Knoten a : Pfad: a, b, c, f, g

Kreise - (2)



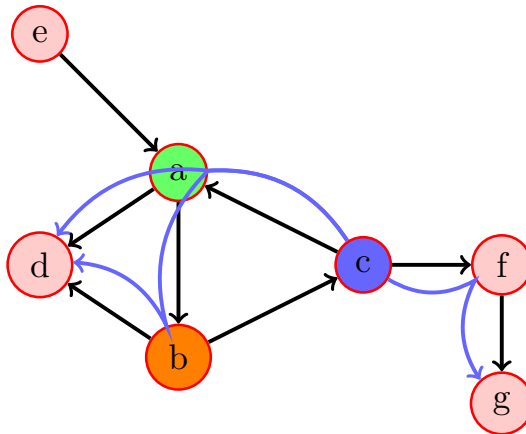
Tiefensuche: Knoten a : Pfad: $a, b, c, f, g \rightarrow \text{neu}(a, g)$

Kreise - (3)

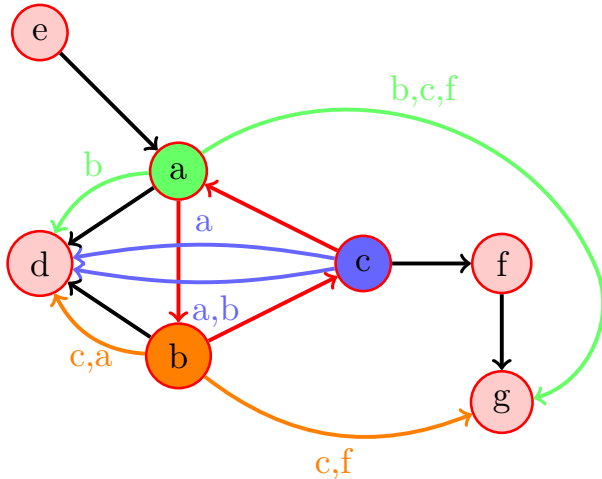


Tiefensuche: Knoten $b \rightarrow$ neu $(b, d), (b, g)$

Kreise - (4)

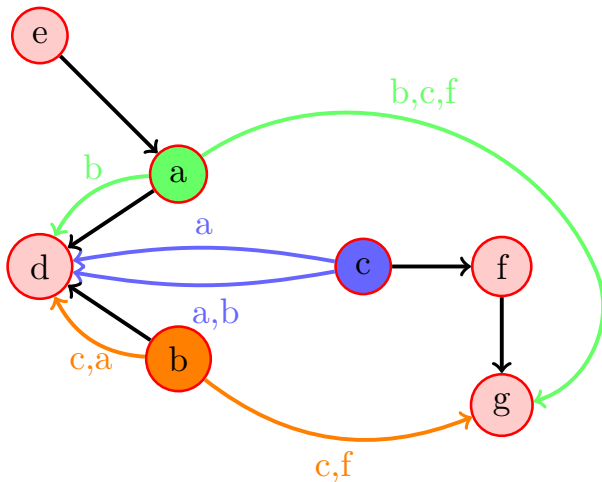


Tiefensuche: Knoten $c \rightarrow$ neu $(c, d), (c, d), (c, g)$



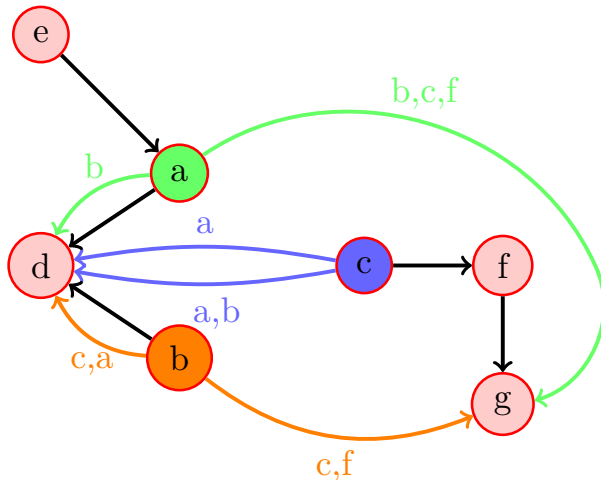
alle neuen Kanten

Kreise - (6)



ohne Kreiskanten

Kreise - (6)



ohne Kreiskanten \Rightarrow zyklusfreier Graph

intro

mcluster

bdc

problem

related

mdc

others

conclusion

ideas

concept

precondition

abstraction

transformation

cycles

full auto

example

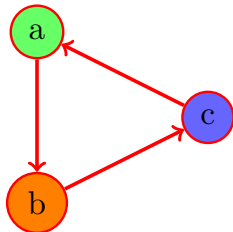
extensions

tests

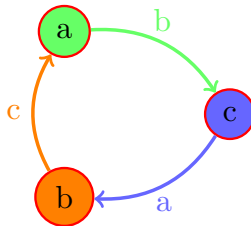
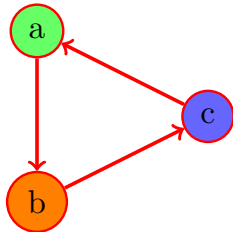
tpch

benchmarks

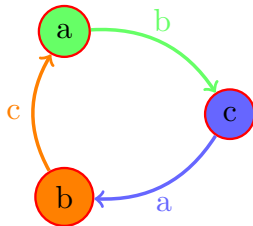
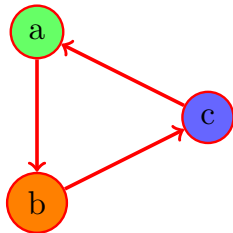
Graph = kompletter Kreis



Graph = kompletter Kreis

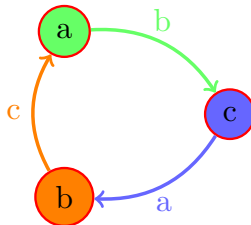
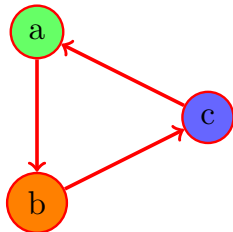


Graph = kompletter Kreis



⇒ spezielle Behandlung

Graph = kompletter Kreis



⇒ spezielle Behandlung → nicht weiter betrachtet

= komplett automatisierte Variante

intro

mcluster
bdc
problem

related

mdc
others
conclusion

ideas

conecpt
precondition
abstraction
transformation
cycles

full auto

example
extensions

tests

tpch
bechmarks

intro

mcluster
bdc
problem

related

mdc
others
conclusion

ideas

conecpt
precondition
abstraction
transformation
cycles
full auto
example
extensions

tests

tpch
bechmarks

= komplett automatisierte Variante
Ablauf:

intro

mcluster
bdc
problem

related

mdc
others
conclusion

ideas

concept
precondition
abstraction
transformation
cycles
full auto
example
extensions

tests

tpch
bechmarks

= komplett automatisierte Variante

Ablauf:

▷ Schema → Graph

intro

mcluster
bdc
problem

related

mdc
others
conclusion

ideas

concept
precondition
abstraction
transformation
cycles

full auto

example
extensions

tests

tpch
benchmarks

= komplett automatisierte Variante

Ablauf:

- ▷ Schema → Graph
- ▷ Dimensionen finden /registrieren

intro

mcluster
bdc
problem

related

mdc
others
conclusion

ideas

concept
precondition
abstraction
transformation
cycles
full auto
example
extensions

tests

tpch
bechmarks

= komplett automatisierte Variante

Ablauf:

- ▷ Schema → Graph
- ▷ Dimensionen finden /registrieren
- ▷ BDC Tabellen erzeugen: von außen nach innen

intro

mcluster
bdc
problem

related

mdc
others
conclusion

ideas

concept
precondition
abstraction
transformation
cycles
full auto
example
extensions

tests

tpch
bechmarks

= komplett automatisierte Variante

Ablauf:

- ▷ Schema → Graph
- ▷ Dimensionen finden /registrieren
- ▷ BDC Tabellen erzeugen: von außen nach innen

Grundvariante - (1)

intro

mcluster
bdc
problem

related

mdc
others
conclusion

ideas

conecpt
precondition
abstraction
transformation
cycles
full auto
example
extensions

tests

tpch
bechmarks

Dimensionen finden/registrieren:

intro

mcluster
bdc
problem

related

mdc
others
conclusion

ideas

conecpt
precondition
abstraction
transformation
cycles
full auto
example
extensions

tests

tpch
bechmarks

Dimensionen finden/registrieren:

▷ Blätter finden

Grundvariante - (1)

intro

mcluster
bdc
problem

related

mdc
others
conclusion

ideas

concept
precondition
abstraction
transformation
cycles
full auto
example
extensions

tests

tpch
bechmarks

Dimensionen finden/registrieren:

- ▷ Blätter finden
- ▷ für jedes Blatt:

intro

mcluster
bdc
problem

related

mdc
others
conclusion

ideas

concept
precondition
abstraction
transformation
cycles

full auto

example
extensions

tests

tpch
benchmarks

Dimensionen finden/registrieren:

- ▷ Blätter finden
- ▷ für jedes Blatt:
 - ◇ Dimension registrieren

intro

mcluster
bdc
problem

related

mdc
others
conclusion

ideas

concept
precondition
abstraction
transformation
cycles
full auto
example
extensions

tests

tpch
bechmarks

Dimensionen finden/registrieren:

- ▷ Blätter finden
- ▷ für jedes Blatt:
 - ◇ Dimension registrieren
 - ◇ BDC-Tabelle erzeugen

Grundvariante - (2)

intro

mcluster
bdc
problem

related

mdc
others
conclusion

ideas

conecpt
precondition
abstraction
transformation
cycles

full auto

example
extensions

tests

tpch
bechmarks

BDC-Tabellen erzeugen:

- ▷ ermittle Dimensionen der abgearbeiteten Nachbartabellen

Grundvariante - (2)

intro

mcluster
bdc
problem

related

mdc
others
conclusion

ideas

concept
precondition
abstraction
transformation
cycles
full auto
example
extensions

tests

tpch
bechmarks

BDC-Tabellen erzeugen:

- ▷ ermittle Dimensionen der abgearbeiteten Nachbartabellen
- ▷ sammle Pfade zu Dimtab mit Δ + Nachbar-Dim Pfad

Grundvariante - (2)

intro

mcluster
bdc
problem

related

mdc
others
conclusion

ideas

concept
precondition
abstraction
transformation
cycles
full auto
example
extensions

tests

tpch
bechmarks

BDC-Tabellen erzeugen:

- ▷ ermittle Dimensionen der abgearbeiteten Nachbartabellen
- ▷ sammle Pfade zu Dimtab mit Δ + Nachbar-Dimfad
- ▷ benötigte *iohits*, aus System ermittelt

Grundvariante - (2)

intro

mcluster
bdc
problem

related

mdc
others
conclusion

ideas

concept
precondition
abstraction
transformation
cycles

full auto

example
extensions

tests

tpch
bechmarks

BDC-Tabellen erzeugen:

- ▷ ermittle Dimensionen der abgearbeiteten Nachbartabellen
- ▷ sammle Pfade zu Dimtab mit Δ + Nachbar-Dimfad
- ▷ benötigte *iobits*, aus System ermittelt
- ▷ *Masken* = *RoundRobin*(*Dims*, *iobits*)

Grundvariante - (2)

intro

mcluster
bdc
problem

related

mdc
others
conclusion

ideas

concept
precondition
abstraction
transformation
cycles
full auto
example
extensions

tests

tpch
bechmarks

BDC-Tabellen erzeugen:

- ▷ ermittle Dimensionen der abgearbeiteten Nachbartabellen
- ▷ sammle Pfade zu Dimtab mit Δ + Nachbar-Dimfad
- ▷ benötigte *iobits*, aus System ermittelt
- ▷ *Masken* = *RoundRobin*(*Dims*, *iobits*)
- ▷ erzeuge BDC-Tabelle

intro

mcluster
bdc
problem

related

mdc
others
conclusion

ideas

concept
precondition
abstraction
transformation
cycles
full auto
example
extensions

tests

tpch
bechmarks

BDC-Tabellen erzeugen:

- ▷ ermittle Dimensionen der abgearbeiteten Nachbartabellen
- ▷ sammle Pfade zu Dimtab mit Δ + Nachbar-Dimfad
- ▷ benötigte *iobits*, aus System ermittelt
- ▷ *Masken* = *RoundRobin*(*Dims*, *iobits*)
- ▷ erzeuge BDC-Tabelle
- ▷ speichere Diminfos

Grundvariante - (2)

intro

mcluster
bdc
problem

related

mdc
others
conclusion

ideas

concept
precondition
abstraction
transformation
cycles
full auto
example
extensions

tests

tpch
benchmarks

BDC-Tabellen erzeugen:

- ▷ ermittle Dimensionen der abgearbeiteten Nachbartabellen
- ▷ sammle Pfade zu Dimtab mit Δ + Nachbar-Dim Pfad
- ▷ benötigte *iobits*, aus System ermittelt
- ▷ *Masken* = *RoundRobin*(*Dims*, *iobits*)
- ▷ erzeuge BDC-Tabelle
- ▷ speichere Diminfos
- ▷ markiere Tabelle als bearbeitet

Grundvariante - (2)

intro

mcluster
bdc
problem

related

mdc
others
conclusion

ideas

concept
precondition
abstraction
transformation
cycles
full auto
example
extensions

tests

tpch
benchmarks

BDC-Tabellen erzeugen:

- ▷ ermittle Dimensionen der abgearbeiteten Nachbartabellen
- ▷ sammle Pfade zu Dimtab mit Δ + Nachbar-Dimfad
- ▷ benötigte *iohits*, aus System ermittelt
- ▷ *Masken* = *RoundRobin(Dims, iohits)*
- ▷ erzeuge BDC-Tabelle
- ▷ speichere Diminfos
- ▷ markiere Tabelle als bearbeitet

intro

mcluster

bdc

problem

related

mdc

others

conclusion

ideas

concept

precondition

abstraction

transformation

cycles

full auto

example

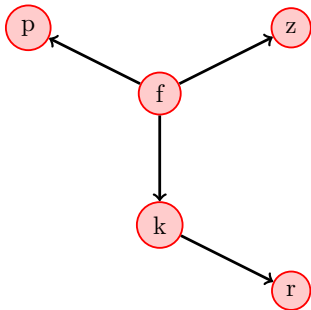
extensions

tests

tpch

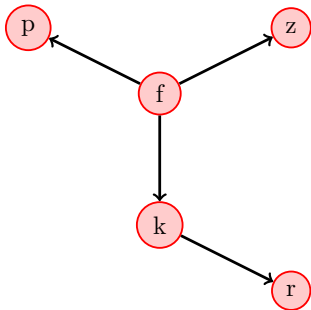
benchmarks

Beispiel (Graph)



Transformierter Schemagraph

Beispiel (Graph)

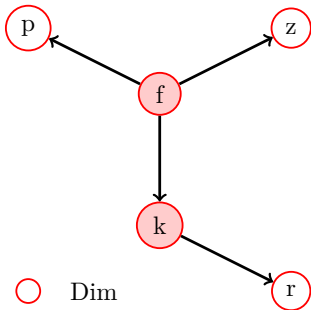


Transformierter Schemagraph

- ▷ zusammenhängend
- ▷ ohne Kreise

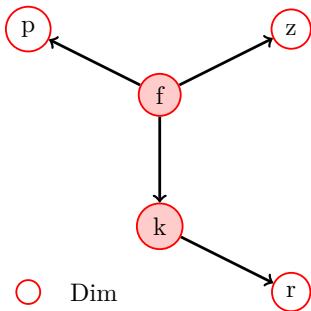
Beispiel - (Dimensionen)

Dimensionen = Blätter



Beispiel - (Dimensionen)

Dimensionen = Blätter
Dim-Registrierung:



intro

mcluster
bdc
problem

related

mdc
others
conclusion

ideas

concept
precondition
abstraction
transformation
cycles
full auto

example

extensions

tests

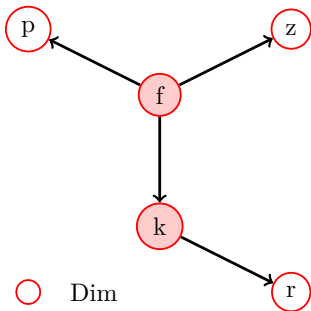
tpch
benchmarks

Beispiel - (Dimensionen)

Dimensionen = Blätter

Dim-Registrierung:

▷ Dimspalte = Primärschlüssel

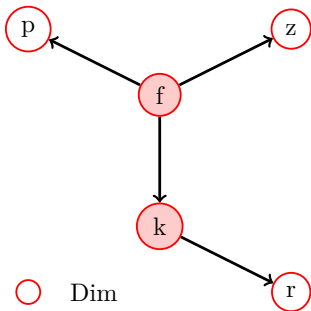


Beispiel - (Dimensionen)

Dimensionen = Blätter

Dim-Registrierung:

- ▷ Dimspalte = Primärschlüssel
- ▷ $cbits = \min\{\lfloor \log_2(count(x)) \rfloor, max_{cbits}\}$
 $\forall x$ Dimtab



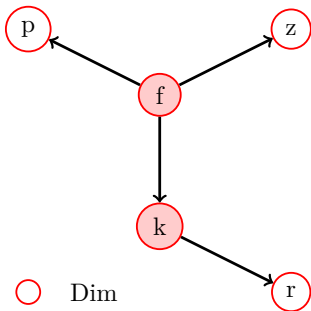
Beispiel - (Dimensionen)

Dimensionen = Blätter

Dim-Registrierung:

- ▷ Dimspalte = Primärschlüssel
- ▷ $cbits = \min\{\lfloor \log_2(count(x)) \rfloor, max_{cbits}\}$
 $\forall x$ Dimtab

BDC-Tabelle erzeugen:



Beispiel - (Dimensionen)

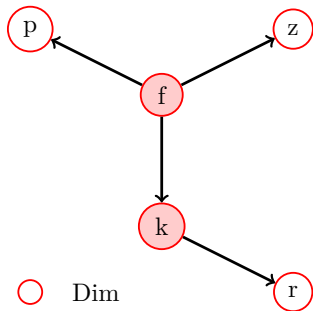
Dimensionen = Blätter

Dim-Registrierung:

- ▷ Dimspalte = Primärschlüssel
- ▷ $cbits = \min\{\lfloor \log_2(count(x)) \rfloor, max_{cbits}\}$
 $\forall x$ Dimtab

BDC-Tabelle erzeugen:

- ▷ $iobits = cbits$



Beispiel - (Dimensionen)

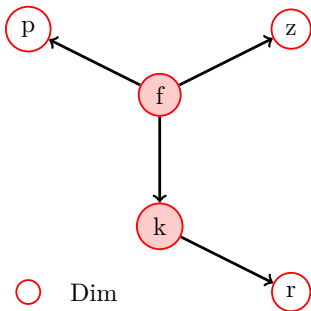
Dimensionen = Blätter

Dim-Registrierung:

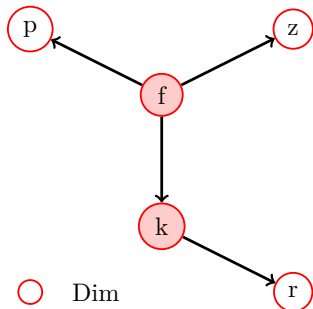
- ▷ Dimspalte = Primärschlüssel
- ▷ $cbits = \min\{\lfloor \log_2(count(x)) \rfloor, max_{cbits}\}$
 $\forall x$ Dimtab

BDC-Tabelle erzeugen:

- ▷ $iobits = cbits$
- ▷ $mask = 1...1, iobit$ viele



Beispiel - (Dimensionen)



Dimensionen = Blätter

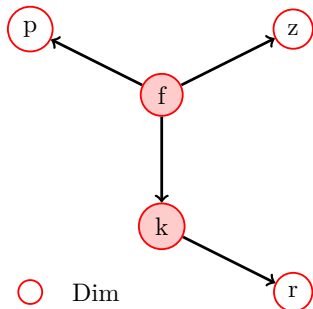
Dim-Registrierung:

- ▷ Dimspalte = Primärschlüssel
- ▷ $cbits = \min\{\lfloor \log_2(count(x)) \rfloor, max_{cbits}\}$
 $\forall x$ Dimtab

BDC-Tabelle erzeugen:

- ▷ $iobits = cbits$
- ▷ $mask = 1...1$, $iobit$ viele
- ▷ $Pfad(x) = Dimspalte$

Beispiel - (Dimensionen)



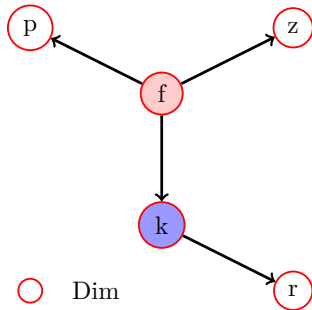
Dimensionen = Blätter

Dim-Registrierung:

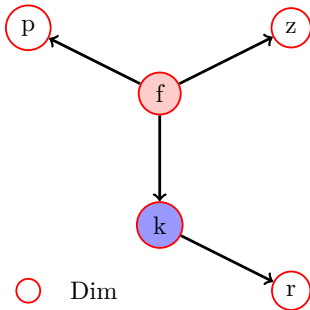
- ▷ Dimspalte = Primärschlüssel
- ▷ $cbits = \min\{\lfloor \log_2(count(x)) \rfloor, max_{cbits}\}$
 $\forall x$ Dimtab

BDC-Tabelle erzeugen:

- ▷ $iobits = cbits$
- ▷ $mask = 1...1$, $iobit$ viele
- ▷ $Pfad(x) =$ Dimspalte
- ▷ Dimtab als abgearbeitet markieren

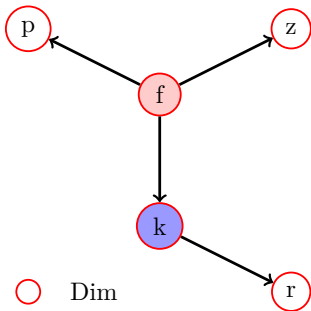
Beispiel - (Tabelle k)

Beispiel - (Tabelle k)



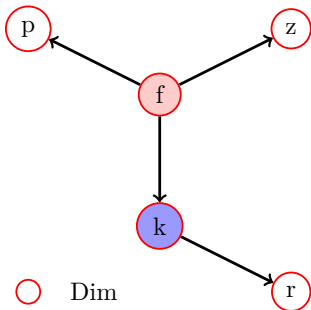
$$\triangleright \text{Pfad}(k) = (k, r) + \text{Pfad}(r)$$

Beispiel - (Tabelle k)



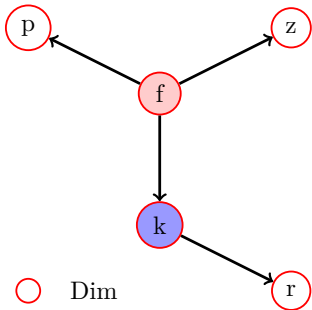
- ▷ $Pfad(k) = (k, r) + Pfad(r)$
- ▷ *iobits* vom System ausgelesen

Beispiel - (Tabelle k)

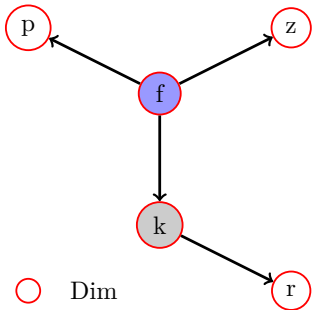


- ▷ $Pfad(k) = (k, r) + Pfad(r)$
- ▷ $iobits$ vom System ausgelesen
- ▷ $mask = RoundRobin(\{r\}, iobits)$

Beispiel - (Tabelle k)



- ▷ $Pfad(k) = (k, r) + Pfad(r)$
- ▷ $iobits$ vom System ausgelesen
- ▷ $mask = RoundRobin(\{r\}, iobits)$
- ▷ k abgearbeitet

Beispiel - (Tabelle f)

intro

mcluster

bdc

problem

related

mdc

others

conclusion

ideas

concept

precondition

abstraction

transformation

cycles

full auto

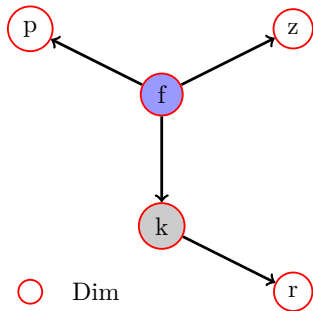
example

extensions

tests

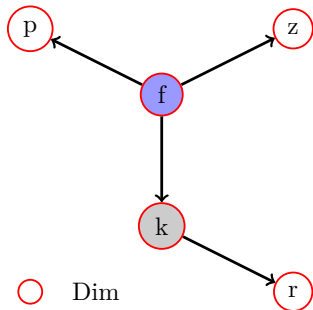
tpch

benchmarks

Beispiel - (Tabelle f)

$$\triangleright P\text{fad}_1(f) = (f, p) + P\text{fad}(p)$$

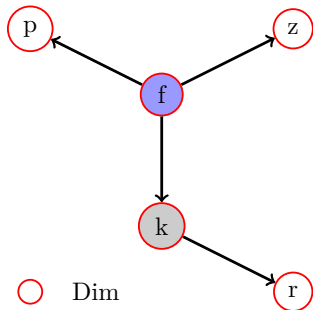
Beispiel - (Tabelle f)



$$\triangleright P\text{fad}_1(f) = (f, p) + P\text{fad}(p)$$

$$\triangleright P\text{fad}_2(f) = (f, z) + P\text{fad}(z)$$

Beispiel - (Tabelle f)

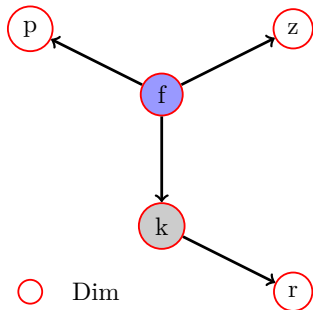


$$\triangleright P\text{fad}_1(f) = (f, p) + P\text{fad}(p)$$

$$\triangleright P\text{fad}_2(f) = (f, z) + P\text{fad}(z)$$

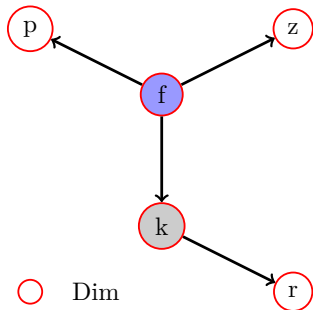
$$\triangleright P\text{fad}_3(f) = (f, k) + P\text{fad}(k)$$

Beispiel - (Tabelle f)



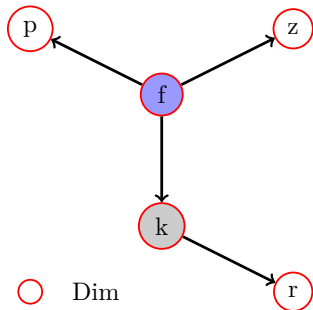
- ▷ $Pfad_1(f) = (f, p) + Pfad(p)$
- ▷ $Pfad_2(f) = (f, z) + Pfad(z)$
- ▷ $Pfad_3(f) = (f, k) + Pfad(k)$
- ▷ *iohits* vom System ausgelesen

Beispiel - (Tabelle f)



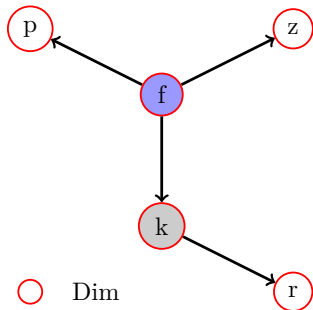
- ▷ $Pfad_1(f) = (f, p) + Pfad(p)$
- ▷ $Pfad_2(f) = (f, z) + Pfad(z)$
- ▷ $Pfad_3(f) = (f, k) + Pfad(k)$
- ▷ $iobits$ vom System ausgelesen
- ▷ $masks =$
 $RoundRobin(\{p, z, k\}, iobits)$

Beispiel - (Tabelle f)



- ▷ $Pfad_1(f) = (f, p) + Pfad(p)$
- ▷ $Pfad_2(f) = (f, z) + Pfad(z)$
- ▷ $Pfad_3(f) = (f, k) + Pfad(k)$
- ▷ $iobits$ vom System ausgelesen
- ▷ $masks =$
 $RoundRobin(\{p, z, k\}, iobits)$
- ▷ f abgearbeitet

Beispiel - (Tabelle f)



- ▷ $Pfad_1(f) = (f, p) + Pfad(p)$
- ▷ $Pfad_2(f) = (f, z) + Pfad(z)$
- ▷ $Pfad_3(f) = (f, k) + Pfad(k)$
- ▷ $iobits$ vom System ausgelesen
- ▷ $masks =$
 $RoundRobin(\{p, z, k\}, iobits)$
- ▷ f abgearbeitet \rightarrow fertig

intro

mcluster
bdc
problem

related

mdc
others
conclusion

ideas

concept
precondition
abstraction
transformation
cycles
full auto
example
extensions

tests

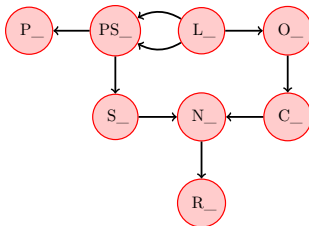
tpch
benchmarks

Beschrieben:

- ▷ Wunschdimensionen

Offen:

- ▷ zu viele Dimensionen
- ▷ Anfrageanalyse vorschalten
- ▷ Schema = Kreis



▷ DB Test für Data-Warehouse Szenarien

intro

mcluster
bdc
problem

related

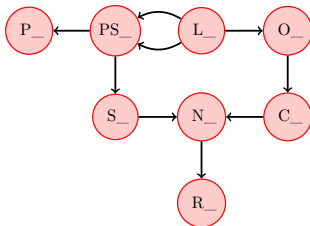
mdc
others
conclusion

ideas

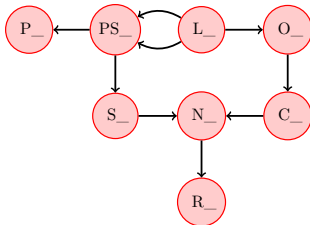
concept
precondition
abstraction
transformation
cycles
full auto
example
extensions

tests

tpch
benchmarks



- ▷ DB Test für Data-Warehouse Szenarien
- ▷ Anfragen
 - ◇ komplex
 - ◇ ohne Vorwissen
 - ◇ große Datenmengen (abhängig von SF)



- ▷ DB Test für Data-Warehouse Szenarien
- ▷ Anfragen
 - ◇ komplex
 - ◇ ohne Vorwissen
 - ◇ große Datenmengen (abhängig von SF)
- ▷ gefundene Dimensionen: R_-, P_-
- ▷ $SF = 1 \rightarrow$ ca. 1 GB

intro

mcluster

bdc

problem

related

mdc

others

conclusion

ideas

concept

precondition

abstraction

transformation

cycles

full auto

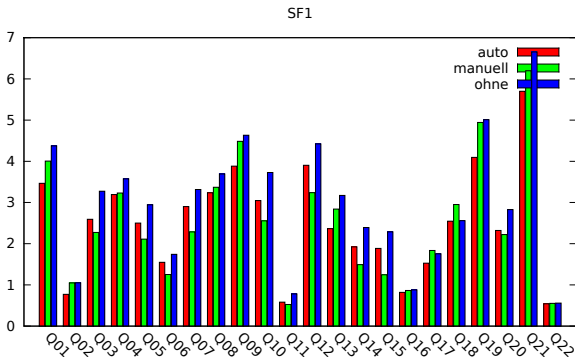
example

extensions

tests

tpch

benchmarks



3 Wdh., pro Query Kaltstart, Anfragen nicht auf Schema optimiert
 Q5 Umsätze der lokalen Anbieter in S_

