Datenbank Implementierungstechniken

mehrdimensionale Dateiorganisation und Zugriffspfade

Max Taube

29. Juni 2015

HTWK Leipzig

Ablauf

Anforderungen

Umsetzung

- kdB-Baum
- mehrdimensionales Hashing
- Grid-File

Anforderungen

Anforderungen

Ziel: Finden von Datensätzen durch Attribute

- Identifikation über mehrere Attribute
- ein Attribut bildet eine Dimension
- Abbildung der Datensätze in mehrdimensionalen Datenraum
- $\hookrightarrow$ mehrdimensionale Indexe
- erhalten der topologischen Struktur
  - benachbarte Objekt auch im Speicher nah

Anforderungen

Match Suche mit Attributen

Partial-Match Suche mit nicht allen Attributen

Range-Match Suche mit Bereich von Attributen

↪ Partial- und Range-Match mit eindimensionalen Indexen sehr aufwändig

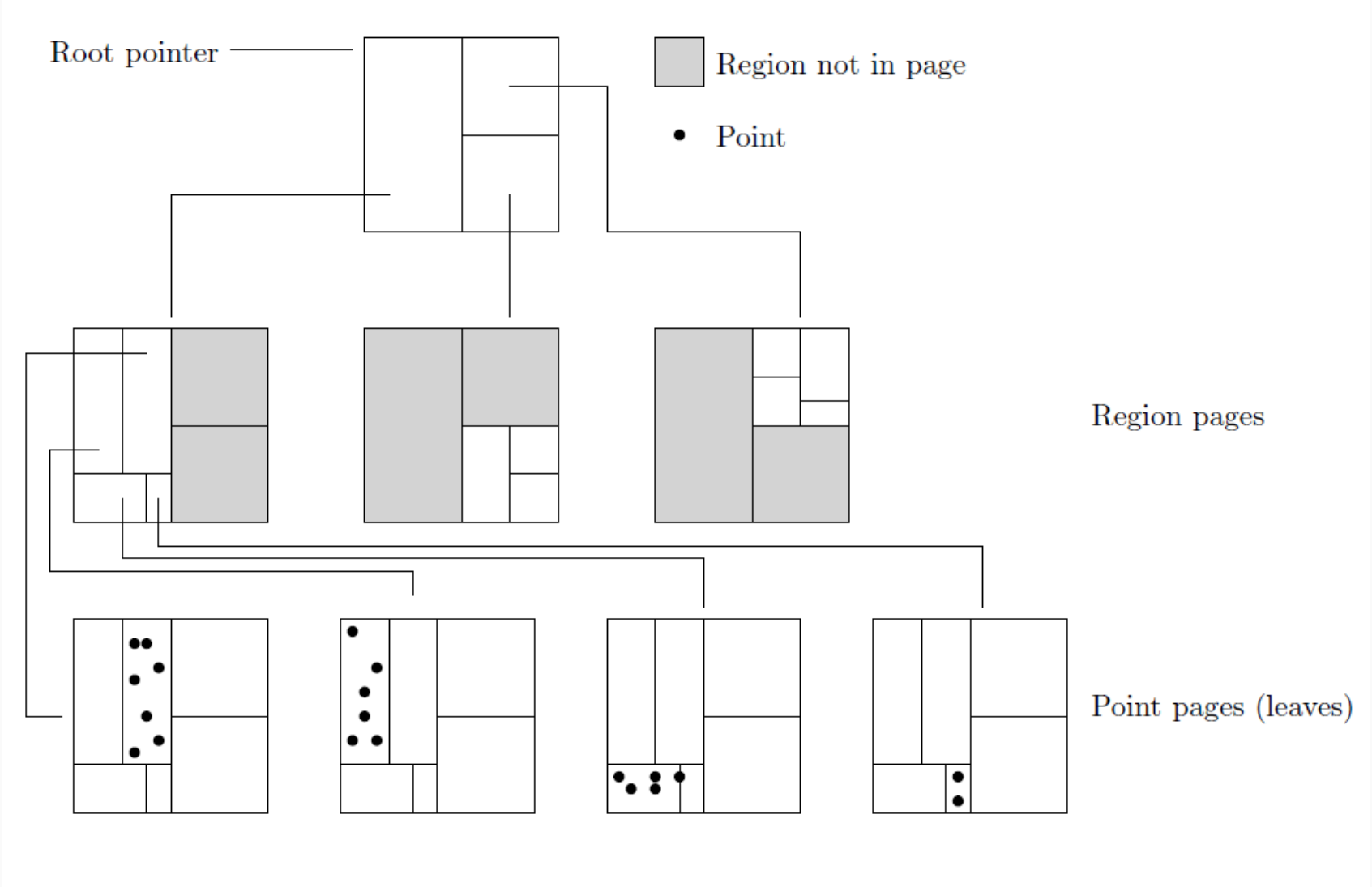
Umsetzung

kdB-Baum

kdB-Baum

- B+-Baum mit k Dimensionen
- Jeder Index-Knoten partitioniert den Suchraum in mehrere Quader
  - Index-Knoten: Menge von Region-Child-Paaren
  - Blatt-Knoten: Menge von Punkt-Referenz-Paaren

kdB-Baum


Root pointer
Region not in page
Point
Region pages
Point pages (leaves)

kdB-Baum

- Reorganisation bei Überläufen
- Balancierung
- Range- und Partial-Match gut möglich

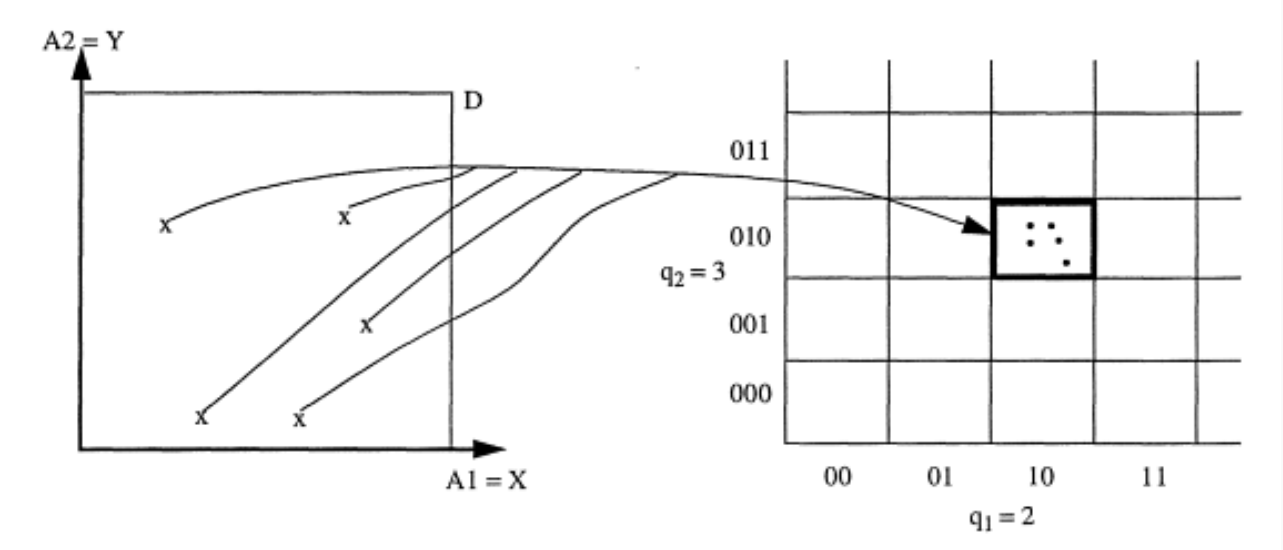
mehrdimensionales Hashing

mehrdimensionales Hashing

Hashfunktion bildet Schlüssel auf Bucket ab

- Hashfunktion mit mehreren Attributen lässt keine Partial- oder Range-Matches zu
- Nutzung von separaten Hashfunktionen für jedes Attribut
- Konkatenation der Teil-Hashes
  - Hashfunktion der Attribute müssen so gewählt sein das gewünschte Gesamtlänge des Hashes erreicht wird (Q Bits für 2^Q Buckets)

mehrdimensionales Hashing



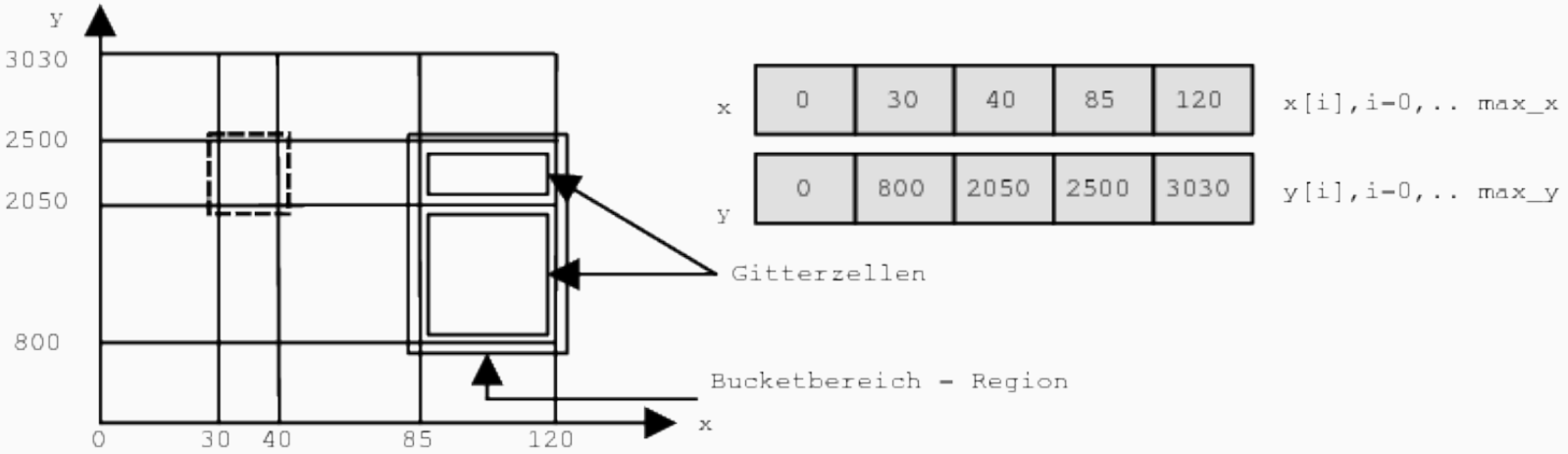
- Partial-Match durch bilden aller Hashes
  - A2 = 010 $\rightarrow$ 01000 01001 01010 01011
- topologische Struktur nicht erhalten
- Range-Match aufwändig
- viele leere Buckets möglich

Grid-File

Grid-File

- Aufteilung des mehrdimensionalen Raums in Quader
- Feld für jede Dimension als Skala
- Grid-Directoy zum Finden des Buckets
  - k-Dimensionales Feld
  - durch Skalen beschrieben

Grid-File



- Partial- und Range-Match gut möglich
- nichtlineares Wachstum
- viele Nullzellen möglich
- topologische Struktur erhalten

Quellen

Thomas Kudraß.
Taschenbuch Datenbanken, 2007.

Erhard Rahm Theo Härder.
Datenbanksysteme: Konzepte und Techniken der Implementierung.
Springer-Verlag, 2001.

Wikipedia.
K-D-B-tree.
en.wikipedia.org/wiki/K-D-B-tree.