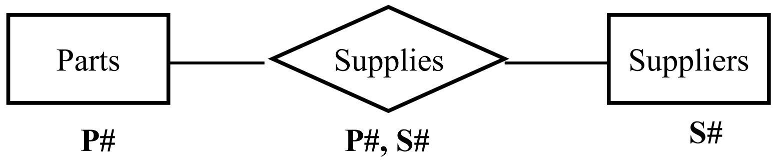
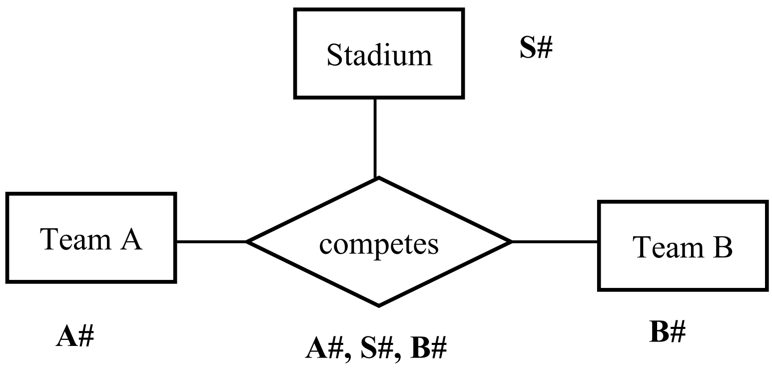
Chapter 10.1: UB-tree for Multidimensional Indexing

Chapter 6.1 Introduction

Note: ***all relational databases are multidimensional: a tuple in a relation with m attributes is considered as a point in m-dimensional space.***





The relation ***competes*** resulting from the relationship is 3-dimensional just considering keys, even more with additional attributes

Geographic Data: weather stations:

(**X, Y, Z, time**, temp, humidity, wind velocity, ...)

Fundamental Problem: ***how to partition multidimensional space for fast search, insertion, deletion?***

UB-tree relies on 3 basic concepts:

area
address
region

Areas and Addresses

Definition: An ***area A*** is a special subspace of the hypercube universe constructed as follows: partition the m-dim cube into 2^m subcubes of equal size and number them $1,2,...,2^m$

1. at level 1 take the first a_1 subcubes
2. at level 2 take the first a_2 subcubes of subcube a_1+1 of level 1
3. at level 3 take the first a_3 subcubes of subcube a_2+1 of level 2

and in general:

k. at level k take the first a_k subcubes of subcube $a_{k-1}+1$ of level k-1

etc.

Definition: The ***address of an area A*** is the sequence of subcube numbers

$a_1.a_2. \ ... \ .a_k$ = alpha (A) = α (A)

according to the preceding definition.

Note: Lexicographic increase of address makes area bigger

Theorem: $\alpha < \beta \Longleftrightarrow \mathbf{Area}(\alpha) \subseteq \mathbf{Area}(\beta)$

Concept of region

Def. If $\alpha < \beta$ **,** then

region $[\alpha : \beta] = \rho\,[\alpha : \beta] :=$ **Area** (β) **- Area**(α)

Note: for an increasing sequence of addresses

$\alpha_1 < \alpha_2 < \alpha_3 < \ ... \ < \alpha_k$

we define corresponding regions

$\rho_1 := (0 : \alpha_1] \quad \rho_2 := (\alpha_1 : \alpha_2] \ ... \quad \rho_k := (\alpha_{k-1} : \alpha_k]$

which can also be represented as the increasing sequence:

$\alpha_1 ; \alpha_2 ; \alpha_3 ; \ ... \ ;\alpha_k$

(which will later be stored in a B-tree index)

Alternative definition: an ***area*** is the set of points on an initial interval of the space filling Z-curve

Examples: areas with addresses

2 3.2.1 2.1.3.1.2

Idea 1: store data in region ρ_j on a disk page $\mathbf{P_j}$

Idea 2: store addresses

α_1 ; α_2 ; α_3 ; ... ;α_k

in a B-tree or B*-tree or Prefix B-tree

P_1, …, P_k are the leaves of the B-tree

Fundamental Question: ***How to split spatial region, if the leaf page of a B-tree must be split?***

Points and Coordinates

point p = $(x_1, x_2, \dots, x_m)$

At a certain resolution, p is a small square (pixel)

Def: Address ***alpha(p)*** = address ot that area, whose last point is p

Def: ***cart(α)*** = Cartesian coordinate of point p with address α

Lemma: cart (alpha (p)) = p

cart (alpha $(x_1, x_2, \dots, x_m)$) = $(x_1, x_2, \dots, x_m)$