



Principles of Distributed and Parallel Database Systems

VF: Information Requirements

Objectives



Objective

Realize how queries are
processed in distributed
databases

VF: Information Requirements

Application Information:

| Attribute affinities

- a measure that indicates how closely related the attributes are
- This is obtained from more primitive usage data

| Attribute usage values

- Given a set of queries $Q = \{q_1, q_2, \dots, q_q\}$ that will run on the relation: $R[A_1, A_2, \dots, A_n]$,

$$use(q_i, A_j) = \begin{cases} 1 & \text{if attribute } A_j \text{ is referenced by query } q_i \\ 0 & \text{otherwise} \end{cases}$$

$use(q_i, \bullet)$ can be defined accordingly

VF – Definition of $use(q_i, A_j)$

| Consider the following 4 queries for relation PROJ:

q_1 : SELECT BUDGET
 FROM PROJ
 WHERE PNO=Value

q_2 : SELECT PNAME,BUDGET
 FROM PROJ

q_3 : SELECT PNAME
 FROM PROJ
 WHERE LOC=Value

q_4 : SELECT SUM(BUDGET)
 FROM PROJ
 WHERE LOC=Value

	A_1	A_2	A_3	A_4
q_1	1	0	1	0
q_2	0	1	1	0
q_3	0	1	0	1
q_4	0	0	1	1

Let A_1 = PNO, A_2 = PNAME, A_3 = BUDGET, A_4 = LOC

Attribute Usage Matrix

VF – Affinity Measure $aff(A_i, A_j)$

| The **attribute affinity measure** between two attributes A_i and A_j of a relation $R[A_1, A_2, \dots, A_n]$ with respect to the set of applications $Q = (q_1, q_2, \dots, q_q)$ is defined as:

$$aff(A_i, A_j) = \sum_{\text{all queries that access } A_i \text{ and } A_j} (\text{query access})$$

$$\text{query access} = \sum_{\text{all sites}} \text{access frequency of a query} * \frac{\text{access}}{\text{execution}}$$

Attribute Affinity Matrix

	A1	A2	A3	A4	A5
A1					
A2	50				
A3	45	48			
A4	1	1	0		
A5	0	0	4	75	

$R_1[K, A_1, A_2, A_3]$

$R_2[K, A_4, A_5]$

Hybrid Fragmentation

