

Notes

Friday, June 7, 2019

2:33 AM

- It is easy to represent a many-to-one relationship in JSON.
but it is difficult to represent a many-to-many relationship.
- SQL++ can't express all queries expressible in datalog.

Size & Cost Estimation

$B(R)$ # of blocks

$T(R)$ # of tuples

$V(R, A)$ # of distinct values of attr A .

o estimate size of a query plan P

• f ← selectivity factor
* worst case

$$f = \frac{1}{V(R, A_1) \cdots V(R, A_n)}$$

$$|\sigma_{A_1=a_1 \text{ and } A_2=a_2 \cdots A_n=a_n}| = \frac{T(R)}{V(R, A_1) \cdots V(R, A_n)}$$

$$|RM(S)| = \frac{|R| * |S|}{\max(V(R, B) * V(S, C))}$$

o I/O cost of selection

- Sequential scan $B(R)$

- Index based selection

 - Clustered $f \cdot B(R)$

 - Unclustered $f \cdot T(R)$