

# Query Cost Estimation

## CS 4750 Database Systems

[A. Silberschatz, H. F. Korth, S. Sudarshan, Database System Concepts, Ch.15]  
[C.M. Ricardo, S.D. Urban, "Databases Illuminated, Ch.13]

## Review 1: SQL and RA

Consider the following schema statements.

```
student(ID, name, dept_name, tot_cred)
takes(ID, course ID, sec id, semester, year, grade)
```

Find IDs and names of all students who have taken more than 3 courses

1. Write SQL query
2. Draw an RA plan

## Review 2: SQL and RA

Consider the following schema statements.

```
emp(empno, ename, job, mgr, hiredate, salary, comm, deptno)
dept(deptno, dname, loc)
```

Find the names of departments where more than three employees are working

1. Write SQL query
2. Draw an RA plan

## Review 2: SQL and RA

```
emp(empno, ename, job, mgr, hiredate, salary, comm, deptno)
dept(deptno, dname, loc)
```

Find the names of departments where more than three employees are working

Can you think of another solution?

## Let's Try: Equivalent RA (1)

Consider the Sailors database

Boats (bid, bname, color)  
Sailors (sid, sname, rating, age)  
Reserves (sid, bid, day)

Find the names of sailors who have reserved a red or green boat

$\pi_{\text{sname}}(\sigma_{\text{color}='red' \text{ OR } \text{color}='green'}(\text{Sailors} \bowtie \text{Reserves} \bowtie \text{Boats}))$

Can you think of an equivalent RA? Then, draw the RA tree.

## Let's Try: Equivalent RA (2)

Consider the Sailors database

Boats (bid, bname, color)  
Sailors (sid, sname, rating, age)  
Reserves (sid, bid, day)

Find the names of sailors who have reserved boat 103

$\pi_{\text{sname}}(\sigma_{\text{bid}=103}(\text{Sailors} \bowtie \text{Reserves}))$

Can you think of an equivalent RA? Then, draw the RA tree.

## Let's Try: Equivalent RA (3)

Consider the Sailors database

Boats (bid, bname, color)  
Sailors (sid, sname, rating, age)  
Reserves (sid, bid, day)

Find the IDs and names of sailors who have not reserved a boat

$\pi_{\text{sid}, \text{sname}}(\text{Sailors} - (\text{Sailors} \bowtie \text{Reserves}))$

Can you think of an equivalent RA? Then, draw the RA tree.

## Let's Try: SELECT

Let's go grocery shopping. Assume we know the following info:

Harris\_Teeter(id, name, category, price)

T(Harris_Teeter)	= 1,000	#of tuples
V(name)	= 900	#of distinct values
V(category)	= 10	#of distinct values
V(price)	= 200	#of distinct values
Range(price)	= [1,50]	range of values

---

```
SELECT name
FROM Harris_Teeter
```

How many tuples do we expect this query to output?

## Let's Try: DISTINCT

Let's go grocery shopping. Assume we know the following info:

Harris\_Teeter(id, name, category, price)

T(Harris_Teeter)	= 1,000	#of tuples
V(name)	= 900	#of distinct values
V(category)	= 10	#of distinct values
V(price)	= 200	#of distinct values
Range(price)	= [1,50)	range of values

```
SELECT DISTINCT name
FROM Harris_Teeter
```

How many tuples do we expect this query to output?

## Let's Try: AGGREGATE

Let's go grocery shopping. Assume we know the following info:

Harris\_Teeter(id, name, category, price)

T(Harris_Teeter)	= 1,000	#of tuples
V(name)	= 900	#of distinct values
V(category)	= 10	#of distinct values
V(price)	= 200	#of distinct values
Range(price)	= [1,50)	range of values

```
SELECT category, COUNT(id)
FROM Harris_Teeter
GROUP BY category
```

How many tuples do we expect this query to output?

## Let's Try: WHERE Value

Let's go grocery shopping. Assume we know the following info:

Harris\_Teeter(id, name, category, price)

T(Harris_Teeter)	= 1,000	#of tuples
V(name)	= 900	#of distinct values
V(category)	= 10	#of distinct values
V(price)	= 200	#of distinct values
Range(price)	= [1,50)	range of values

```
SELECT id, name
FROM Harris_Teeter
WHERE id = 45
```

How many tuples do we expect this query to output?

## Let's Try: WHERE Value

Let's go grocery shopping. Assume we know the following info:

Harris\_Teeter(id, name, category, price)

T(Harris_Teeter)	= 1,000	#of tuples
V(name)	= 900	#of distinct values
V(category)	= 10	#of distinct values
V(price)	= 200	#of distinct values
Range(price)	= [1,50)	range of values

```
SELECT *
FROM Harris_Teeter
WHERE name = 'Milk'
```

How many tuples do we expect this query to output?

## Let's Try: WHERE Range

Let's go grocery shopping. Assume we know the following info:

Harris\_Teeter(id, name, category, price)

T(Harris_Teeter)	= 1,000	#of tuples
V(name)	= 900	#of distinct values
V(category)	= 10	#of distinct values
V(price)	= 200	#of distinct values
Range(price)	= [1,50)	range of values

```
SELECT *
FROM Harris_Teeter
WHERE price < 20
```

How many tuples do we expect this query to output?

## Let's Try: AND

Let's go grocery shopping. Assume we know the following info:

Harris\_Teeter(id, name, category, price)

T(Harris_Teeter)	= 1,000	#of tuples
V(name)	= 900	#of distinct values
V(category)	= 10	#of distinct values
V(price)	= 200	#of distinct values
Range(price)	= [1,50)	range of values

```
SELECT *
FROM Harris_Teeter
WHERE name='Milk' AND
category='meat'
```

## Let's Try: OR

Let's go grocery shopping. Assume we know the following info:

Harris\_Teeter(id, name, category, price)

T(Harris_Teeter)	= 1,000	#of tuples
V(name)	= 900	#of distinct values
V(category)	= 10	#of distinct values
V(price)	= 200	#of distinct values
Range(price)	= [1,50)	range of values

```
SELECT *
FROM Harris_Teeter
WHERE name='Milk' OR
category='meat'
```

## Let's Try: Cartesian Product

Let's go grocery shopping. Assume we know the following info:

Harris\_Teeter(id, name, category, price) | DBville(id, dname, shelf, cost)

T(Harris_Teeter)	= 1,000	T(DBville)	= 2,000
V(name)	= 900	V(dname)	= 1,900
V(category)	= 10	V(shelf)	= 12
V(price)	= 200	V(cost)	= 500
Range(price)	= [1,50)		

```
SELECT *
FROM Harris_Teeter, DBville
```

How many tuples do we expect this query to output?

## Let's Try: JOIN

Let's go grocery shopping. Assume we know the following info:

Harris_Teeter( <u>id</u> ,name,category,price)	DBville( <u>id</u> ,dname,shelf,cost)
T(Harris_Teeter) = 1,000	T(DBville) = 2,000
V(name) = 900	V(dname) = 1,900
V(category) = 10	V(shelf) = 12
V(price) = 200	V(cost) = 500
Range(price) = [1,50)	

```
SELECT *
FROM Harris_Teeter H
NATURAL JOIN DBville D
```

How many tuples do we expect this query to output?

## Let's Try: JOIN on Attr

Let's go grocery shopping. Assume we know the following info:

Harris_Teeter( <u>id</u> ,name,category,price)	DBville( <u>id</u> ,dname,shelf,cost)
T(Harris_Teeter) = 1,000	T(DBville) = 2,000
V(name) = 900	V(dname) = 1,900
V(category) = 10	V(shelf) = 12
V(price) = 200	V(cost) = 500
Range(price) = [1,50)	

```
SELECT *
FROM Harris_Teeter H, DBville D
WHERE H.name = D.dname
```

How many tuples do we expect this query to output?