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]

Fall 2025 – University of Virginia

© Praphamontripong

© Praphamontripong

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

Fall 2025 – University of Virginia © Praphamont

Review 1: SQL and RA

Consider the following schema statements.

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

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

- 1. Write SQL query
- 2. Draw an RA plan

Fall 2025 – University of Virginia ©

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?

Fall 2025 – University of Virginia

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

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

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

Fall 2025 - University of Virginia

© Praphamontripong

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

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

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

Fall 2025 – University of Virginia

© Pranhamontrinong

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.

Fall 2025 - University of Virginia

© Praphamontripong

Let's Try: SELECT

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

Harris_Teeter(<u>id</u>, 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?

Fall 2025 – University of Virginia

Let's Try: DISTINCT

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

Harris_Teeter(id, name, category, price)

 $\begin{array}{lll} \text{T(Harris_Teeter)} &=& 1,000 & \text{\#of tuples} \\ \text{V(name)} &=& 900 & \text{\#of distinct values} \\ \text{V(category)} &=& 10 & \text{\#of distinct values} \\ \text{V(price)} &=& 200 & \text{\#of distinct values} \\ \text{Range(price)} &=& [1,50) & \text{range of values} \\ \end{array}$

SELECT DISTINCT name FROM Harris_Teeter

How many tuples do we expect this query to output?

Fall 2025 - University of Virginia

© Praphamontripong

Let's Try: WHERE Value

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

 ${\tt Harris_Teeter}(\underline{\tt id},\ {\tt name},\ {\tt category},\ {\tt price})$

 $\begin{array}{lll} \text{T(Harris_Teeter)} &=& 1,000 & \text{\#of tuples} \\ \text{V(name)} &=& 900 & \text{\#of distinct values} \\ \text{V(category)} &=& 10 & \text{\#of distinct values} \\ \text{V(price)} &=& 200 & \text{\#of distinct values} \\ \text{Range(price)} &=& [1,50) & \text{range of values} \\ \end{array}$

SELECT id, name FROM Harris_Teeter WHERE id = 45

How many tuples do we expect this query to output?

Fall 2025 – University of Virginia

© Pranhamontrinong

Let's Try: AGGREGATE

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

Harris_Teeter(id, name, category, price)

 $\begin{array}{lll} \textbf{T}(\texttt{Harris_Teeter}) &=& 1,000 & \text{\#of tuples} \\ \textbf{V}(\texttt{name}) &=& 900 & \text{\#of distinct values} \\ \textbf{V}(\texttt{category}) &=& 10 & \text{\#of distinct values} \\ \textbf{V}(\texttt{price}) &=& 200 & \text{\#of distinct values} \\ \textbf{Range}(\texttt{price}) &=& [1,50) & \text{range of values} \\ \end{array}$

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

How many tuples do we expect this guery to output?

Fall 2025 - University of Virginia

© Praphamontripong

Let's Try: WHERE Value

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

Harris_Teeter(id, name, category, price)

 $\begin{array}{lll} T(Harris_Teeter) &=& 1,000 & \#of \ tiples \\ V(name) &=& 900 & \#of \ distinct \ values \\ V(category) &=& 10 & \#of \ distinct \ values \\ V(price) &=& 200 & \#of \ distinct \ values \\ Range(price) &=& [1,50) & range \ of \ values \\ \end{array}$

SELECT *
FROM Harris Teeter

WHERE name = 'Milk'

How many tuples do we expect this query to output?

Fall 2025 – University of Virginia

Let's Try: WHERE Range

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

Harris_Teeter(id, name, category, price)

 $\begin{array}{lll} \text{T(Harris_Teeter)} &=& 1,000 & \text{\#of tuples} \\ \text{V(name)} &=& 900 & \text{\#of distinct values} \\ \text{V(category)} &=& 10 & \text{\#of distinct values} \\ \text{V(price)} &=& 200 & \text{\#of distinct values} \\ \text{Range(price)} &=& [1,50) & \text{range of values} \\ \end{array}$

SELECT *

FROM Harris_Teeter WHERE price < 20

How many tuples do we expect this query to output?

Fall 2025 - University of Virginia

© Praphamontripong

Let's Try: OR

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

Harris_Teeter(id, name, category, price)

 $\begin{array}{lll} \text{T(Harris_Teeter)} &=& 1,000 & \text{\#of tuples} \\ \text{V(name)} &=& 900 & \text{\#of distinct values} \\ \text{V(category)} &=& 10 & \text{\#of distinct values} \\ \text{V(price)} &=& 200 & \text{\#of distinct values} \\ \text{Range(price)} &=& [1,50) & \text{range of values} \\ \end{array}$

SELECT *

FROM Harris_Teeter
WHERE name='Milk' OR
category='meat'

Fall 2025 – University of Virginia

@ Pranhamontrinong

Let's Try: AND

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

Harris_Teeter(<u>id</u>, 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'

Fall 2025 - University of Virginia

© Praphamontripong

Let's Try: Cartesian Product

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

Harris_Teeter(id, name, category, price)
T(Harris_Teeter) = 1,000
V(name) = 900
V(category) = 10
V(price) = 200

DBville(<u>id</u>,dname,shelf,cost)

T(DBville) = 2,000
V(dname) = 1,900
V(shelf) = 12
V(cost) = 500

SELECT *

Range(price)

FROM Harris Teeter, DBville

= [1,50)

How many tuples do we expect this query to output?

Fall 2025 – University of Virginia

Let's Try: JOIN

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

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

```
SELECT *
FROM Harris Teeter H
NATURAL JOIN DBville D
```

How many tuples do we expect this query to output?

```
Fall 2025 - University of Virginia
                                                                     © Praphamontripong
```

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(id,dname,shelf,cost)
 T(Harris Teeter) = 1,000
                                       T(DBville)
                                                   = 2,000
 V(name)
                                       V(dname)
                                                   = 1,900
                                                   = 12
                 = 10
                                       V(shelf)
 V(category)
 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?
```

© Praphamontripong

Fall 2025 - University of Virginia