CSC343 Worksheet 2 Solution

June 13, 2020

1. **Exercise 2.4.1:**

   a) $\sigma_{speed \geq 3.0}$(Movies)

   Models 1005, 1006, 1013 have speed greater than 3.0

model	*speed*	*ram*	*hd*	*price*
1001	2.66	1024	250	2114
1002	2.10	512	250	995
1003	1.42	512	80	478
1004	2.80	1024	250	649
1005	3.20	512	250	630
1006	3.20	1024	320	1049
1007	2.20	1024	200	510
1008	2.20	2048	250	770
1009	2.00	1024	250	650
1010	2.80	2048	300	770
1011	1.86	2048	160	959
1012	2.80	1024	160	649
1013	3.06	512	80	529

Correct Solution:

Relational Algebra:

$\pi_{model}(\sigma_{speed \geq 3.0}(\text{Movies}))$

Query Result:

model
1005
1006
1013

Models 1005, 1006, 1013 have speed greater than 3.0

Notes:

- Select
  - Is indicated by σ
  - **Syntax:** σ_{QUERY}SCHEMA_NAME
  - e.g $\sigma_{length \geq 100 \textbf{ AND } studioName=\text{'}Fox\text{'}}(\text{Movies})$

Relation - Movies

title	*year*	*length*	*inColor*	*studioName*	*producerC#*
Star Wars	1977	124	sciFi	Fox	12345
Galaxy Quest	1999	104	comedy	DreamWorks	67890

b) $\pi_{maker}(\sigma_{hd \geq 100}(\text{Product} \bowtie \text{Laptop}))$

Makers A, E, F, G make laptops with hard-disk of at least 100GB.

maker	*model*	*type*
A	1001	pc
A	1002	pc
A	1003	pc
A	2004	laptop
A	2005	laptop
A	2006	laptop
B	1004	pc
B	1005	pc
B	1006	pc
B	2007	laptop
C	1007	pc
D	1008	pc
D	1009	pc
D	1010	pc
D	3004	printer
D	3005	printer
E	1011	pc
E	1012	pc
E	1013	pc
E	2001	laptop
E	2002	laptop
E	2003	laptop
E	3001	printer
E	3002	printer
E	3003	printer
F	2008	laptop
F	2009	laptop
G	2010	laptop
H	3006	printer
H	3007	printer

Figure 2.20: Sample data for Product

model	*speed*	*ram*	*hd*	*screen*	*price*
2001	2.00	2048	240	20.1	3673
2002	1.73	1024	80	17.0	949
2003	1.80	512	60	15.4	549
2004	2.00	512	60	13.3	1150
2005	2.16	1024	120	17.0	2500
2006	2.00	2048	80	15.4	1700
2007	1.83	1024	120	13.3	1429
2008	1.60	1024	100	15.4	900
2009	1.60	512	80	14.1	680
2010	2.00	2048	160	15.4	2300

(b) Sample data for relation Laptop

Correct Solution:

Relational Algebra:

$\pi_{maker}(\sigma_{hd \geq 100}(\text{Product} \bowtie \text{Laptop}))$

Query Result:

maker
A
E
F
G

Makers A, E, F, G make laptops with hard-disk of at least 100GB.

Notes:

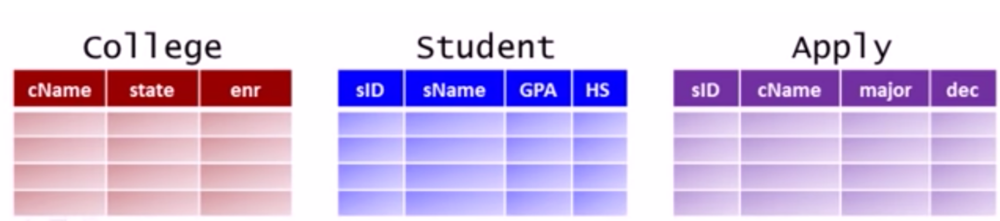
- Project
  - **Syntax:** $\pi_{A_1, A_2, \cdots, A_n}(\text{Rel})$
    - $A_1, \cdots, A_n$ represents attributes
  - Picks certain columns
  - e.g

    What are the titles and years of movies made by Fox that are at least 100 minutes long?

$$\pi_{title,year}(\sigma_{length \geq 100 \textbf{ AND } studioName = \text{'}\textbf{Fox}\text{'}})(\text{Movies})$$

- Cross-Product / Cartesian Product
  - Combines two relations
  - **Syntax:** Relation 1 $\times$ Relation 2
  - e.g. Names and GPAs of students with $HS > 1000$ who applied to CS and were rejected

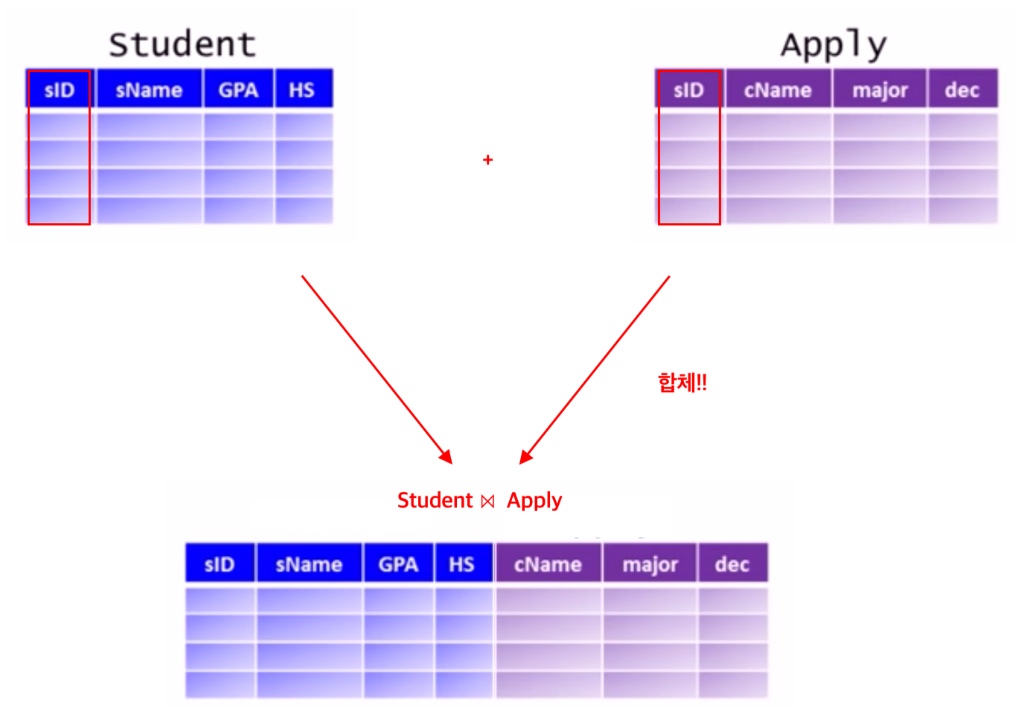
    $\pi_{sName,GPA}(\sigma_{Student.sID = Apply.sID \textbf{ AND } HS > 1000 \textbf{ AND } major = \text{'cs'} \textbf{ AND } dec = \text{'R'}})(\text{Student} \times \text{Apply})$



- Natural Join
  - Enforce equality on all attributes with the same name
  - Eliminiate one copy of duplicate attributes
  - Is symbolized by $\bowtie$
  - **Syntax:** Relation 1 $\bowtie$ Relation 2
  - e.g.

    Names and GPAs of students with $HS > 1000$ who applied to CS and were rejected.

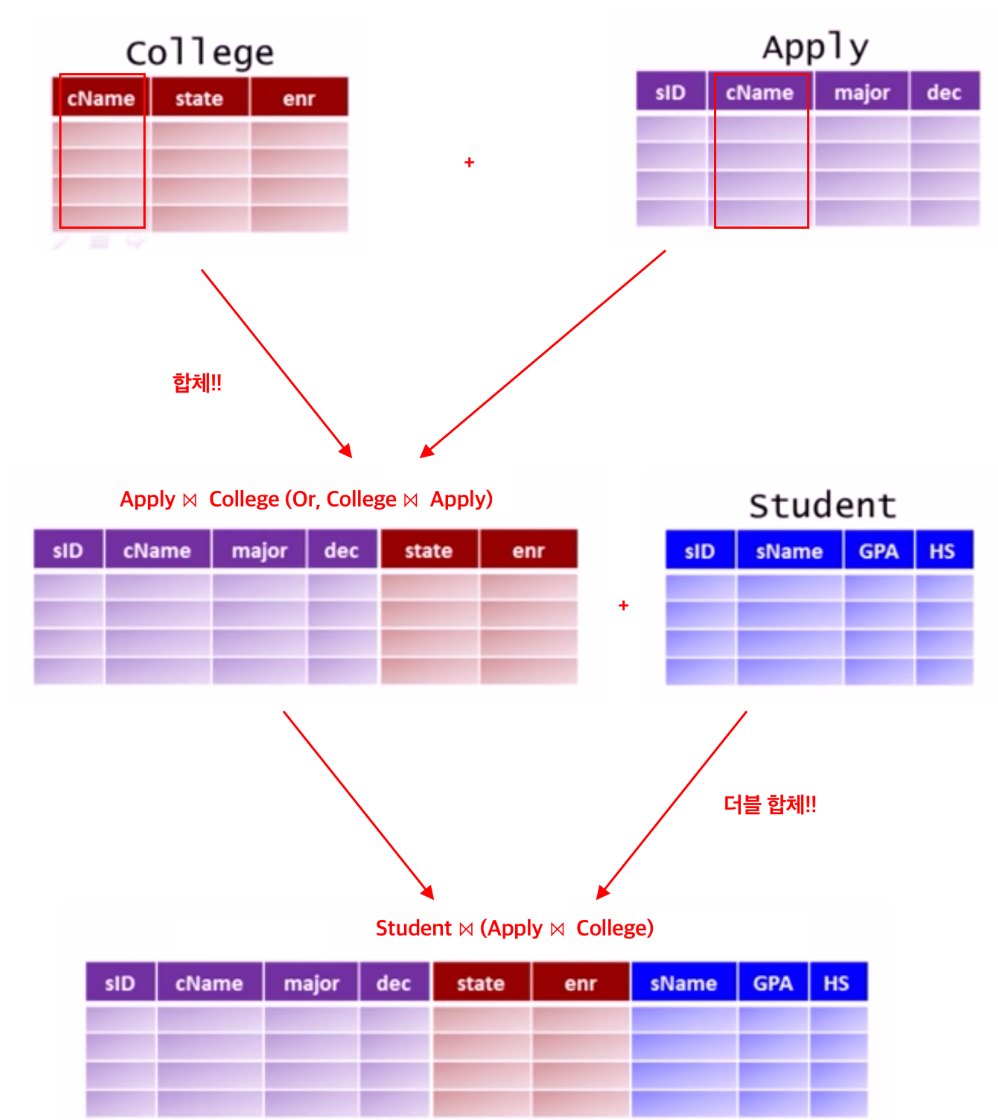
    $$\pi_{sName,GPA}(\sigma_{HS>1000 \textbf{ AND } major=`cs' \textbf{ AND } dec=`R'}(\text{Student} \bowtie \text{Apply}))$$

  - e.g.2.

    Names and GPAs of students with $HS > 1000$ who applied to CS at college with $enr > 20,000$ and were rejected

    $$\pi_{sName,GPA}($$
    $$\sigma_{HS>1000 \textbf{ AND } enr>20000 \textbf{ AND } major=`cs' \textbf{ AND } dec=`R'}(\text{Student} \bowtie (\text{Apply} \bowtie \text{College}))$$
    $$)$$



- Union Operator
  - **Syntax** $\mathrm{R} \cup \mathrm{S}$
  - Is the set of elements that are in R or S or both.
  - An element appears only once in the union even if it is present in both R and S.
  - Is like **UNION** keyword in SQL
  - e.g.

    List of college and student names

$$\pi_{cName}(\text{College}) \cup \pi_{sName}(\text{Student})$$

- Difference Operator

- **Syntax:** $R - S$
- Is also called the *difference* of R and S
- is the set of elements that are in R but not in S.
- Is like **EXCEPT** keyword in SQL
- e.g.

  IDs and names of students who didn't apply anywhere

$$\pi_{sID}(\text{Student}) - \pi_{sID}(\text{Apply})$$

- Intersection Operator
  - **Syntax:** $R \cap S$
  - Is also canned the *intersection* of R and S
  - Is the set of elements that are in both R and S
  - e.g.

    Names that are both a college name and a student name

$$\pi_{cName}(\text{College}) - \pi_{sName}(\text{Student})$$

c)

$$\begin{aligned}&\pi_{model,price}(\sigma_{maker=\text{`B'}}(\\&\text{Product} \bowtie (\pi_{model,price}(\text{Laptop}) \cup \pi_{model,price}(\text{PC}) \cup \pi_{model,price}(\text{Printer})\\&))\end{aligned} \quad (1)$$

The price and model number of all products made by manufacturer B are

1. model 1004, price 649
2. model 1005, price 630
3. model 1006, price 1049
4. model 2007, price 1429

maker	*model*	*type*
A	1001	pc
A	1002	pc
A	1003	pc
A	2004	laptop
A	2005	laptop
A	2006	laptop
B	1004	pc
B	1005	pc
B	1006	pc
B	2007	laptop
C	1007	pc
D	1008	pc
D	1009	pc
D	1010	pc
D	3004	printer
D	3005	printer
E	1011	pc
E	1012	pc
E	1013	pc
E	2001	laptop
E	2002	laptop
E	2003	laptop
E	3001	printer
E	3002	printer
E	3003	printer
F	2008	laptop
F	2009	laptop
G	2010	laptop
H	3006	printer
H	3007	printer

Figure 2.20: Sample data for Product

model	*speed*	*ram*	*hd*	*screen*	*price*
2001	2.00	2048	240	20.1	3673
2002	1.73	1024	80	17.0	949
2003	1.80	512	60	15.4	549
2004	2.00	512	60	13.3	1150
2005	2.16	1024	120	17.0	2500
2006	2.00	2048	80	15.4	1700
2007	1.83	1024	120	13.3	1429
2008	1.60	1024	100	15.4	900
2009	1.60	512	80	14.1	680
2010	2.00	2048	160	15.4	2300

(b) Sample data for relation Laptop

model	*speed*	*ram*	*hd*	*price*
1001	2.66	1024	250	2114
1002	2.10	512	250	995
1003	1.42	512	80	478
1004	2.80	1024	250	649
1005	3.20	512	250	630
1006	3.20	1024	320	1049
1007	2.20	1024	200	510
1008	2.20	2048	250	770
1009	2.00	1024	250	650
1010	2.80	2048	300	770
1011	1.86	2048	160	959
1012	2.80	1024	160	649
1013	3.06	512	80	529

(a) Sample data for relation PC

Correct Solution:

Relational Algebra:

$$\pi_{model,price}(\sigma_{maker='B'}(\text{Product} \bowtie (\pi_{model,price}(\text{Laptop}) \cup \pi_{model,price}(\text{PC}) \cup \pi_{model,price}(\text{Printer}))) \qquad (2)$$

Query Result:

model	price
1004	649
1005	630
1006	1049
2007	1429

The price and model number of all products made by manufacturer B are

1. model 1004, price 649
2. model 1005, price 630
3. model 1006, price 1049
4. model 2007, price 1429

d) $\pi_{model}(\sigma_{color=\text{true } \textbf{AND } type=\text{`laser'}}(\text{Printer}))$

Model 3003, and 3007 are color laster printers

model	*color*	*type*	*price*
3001	true	ink-jet	99
3002	false	laser	239
3003	true	laser	899
3004	true	ink-jet	120
3005	false	laser	120
3006	true	ink-jet	100
3007	true	laser	200

(c) Sample data for relation Printer

Correct Solution:

Relational Algebra:

$\pi_{model}(\sigma_{color=\text{true } \textbf{AND } type=\text{`laser'}}(\text{Printer}))$

Query Result:

model
3003
3007

Model 3003, and 3007 are color laster printers

e) $\pi_{maker}(\text{Product} \bowtie (\pi_{model}(\text{Laptops}) - \pi_{model}(\text{PC})))$

Manufacturers F and G produce laptops but not PCs

maker	model	type
A	1001	pc
A	1002	pc
A	1003	pc
A	2004	laptop
A	2005	laptop
A	2006	laptop
B	1004	pc
B	1005	pc
B	1006	pc
B	2007	laptop
C	1007	pc
D	1008	pc
D	1009	pc
D	1010	pc
D	3004	printer
D	3005	printer
E	1011	pc
E	1012	pc
E	1013	pc
E	2001	laptop
E	2002	laptop
E	2003	laptop
E	3001	printer
E	3002	printer
E	3003	printer
F	2008	laptop
F	2009	laptop
G	2010	laptop
H	3006	printer
H	3007	printer

Figure 2.20: Sample data for Product

model	speed	ram	hd	screen	price
2001	2.00	2048	240	20.1	3673
2002	1.73	1024	80	17.0	949
2003	1.80	512	60	15.4	549
2004	2.00	512	60	13.3	1150
2005	2.16	1024	120	17.0	2500
2006	2.00	2048	80	15.4	1700
2007	1.83	1024	120	13.3	1429
2008	1.60	1024	100	15.4	900
2009	1.60	512	80	14.1	680
2010	2.00	2048	160	15.4	2300

(b) Sample data for relation Laptop

model	speed	ram	hd	price
1001	2.66	1024	250	2114
1002	2.10	512	250	995
1003	1.42	512	80	478
1004	2.80	1024	250	649
1005	3.20	512	250	630
1006	3.20	1024	320	1049
1007	2.20	1024	200	510
1008	2.20	2048	250	770
1009	2.00	1024	250	650
1010	2.80	2048	300	770
1011	1.86	2048	160	959
1012	2.80	1024	160	649
1013	3.06	512	80	529

(a) Sample data for relation PC

Correct Solution:

Relational Algebra:

$$\pi_{maker}(\sigma_{type=\text{'laptop'}\ \textbf{AND}\ type<>\text{'PC'}}(\text{Product}))$$

Query Result:

maker
F
G

Manufacturers F and G produce laptops but not PCs

Notes:

- ‘<>’ Means ‘NOT EQUAL’ in relational algebra
- Relational algebra inclues <u>six</u> comparison operators $(=, <>, <, >, \geq, \leq)$ [1]
- Relational projection (i.e. π) <u>always</u> return distinct tuples [2]

<u>Reference:</u>

1) Radboud University: IS0 - Relational Languages, link
2) Stack Overflow: Selecting DISTINCT rows in relational algebra, link

f) $\pi_{hd}(\sigma_{hd=hd2}(\pi_{hd}(PC) \times \rho_{\pi_{hd}(PC)(hd2)}(\pi_{hd}(PC))))$

Query Result:

hd
250
80
160

<u>Correct Solution:</u>

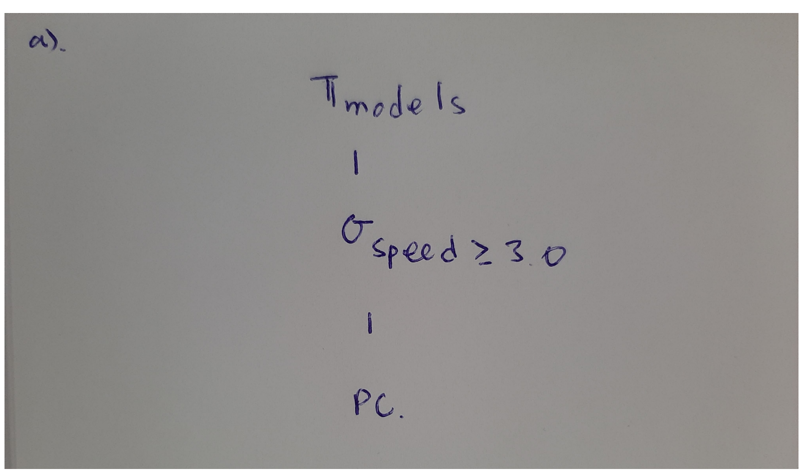
Relational Algebra:

$\pi_{hd}(\sigma_{hd=hd2}(\pi_{hd}(PC) \times \pi_{hd2}(\rho_{hd \rightarrow hd2}(PC))))$

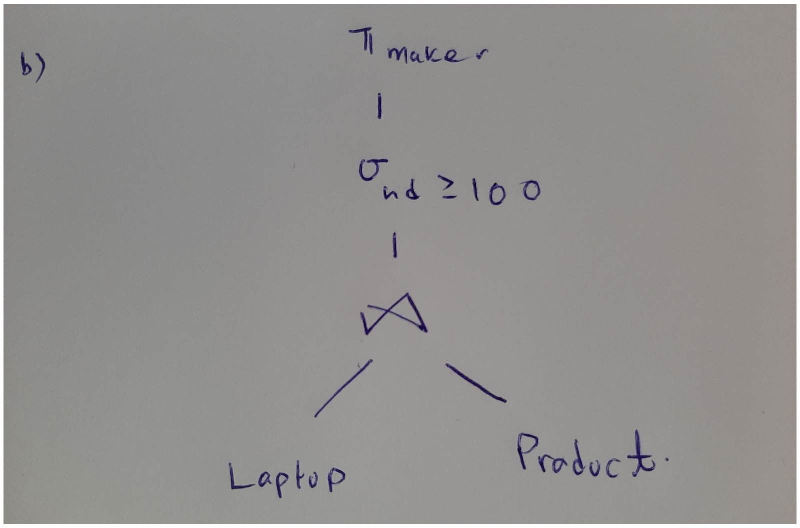
Query Result:

hd
250
80
160

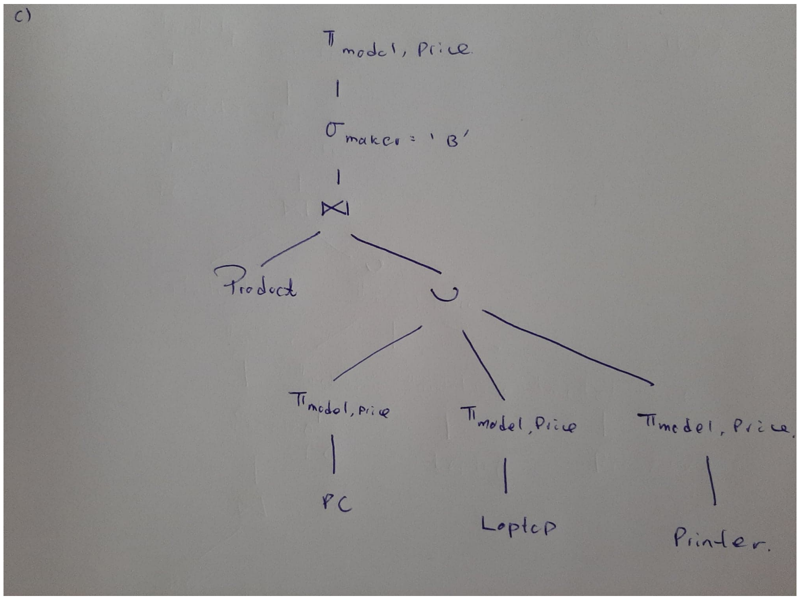
2. a) **Answer:**



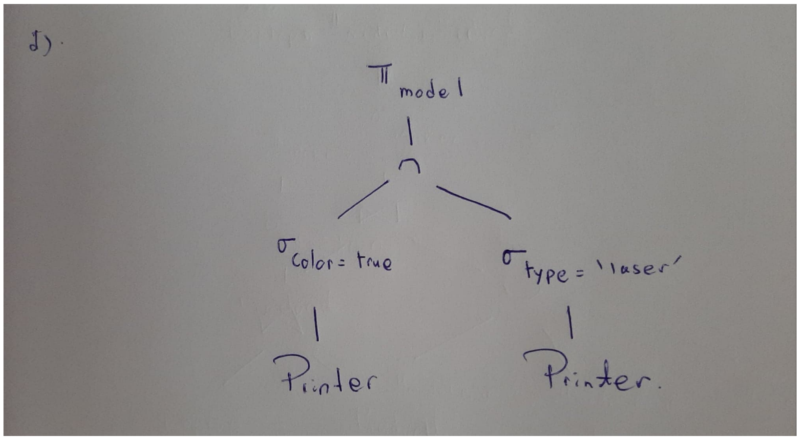
b) **Answer:**



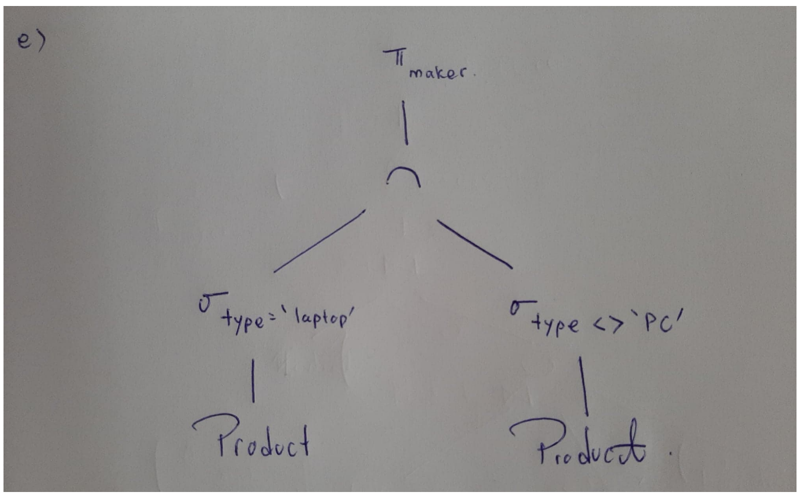
c) **Answer:**



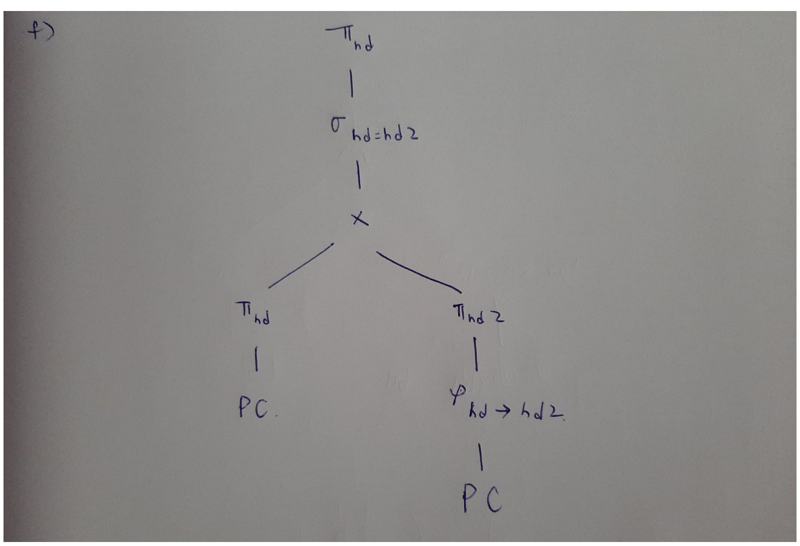
d) **Answer:**



e) **Answer:**



f) **Answer:**



3. a) **Relational Algebra:**

$\pi_{class,countries}(\sigma_{bore \geq 16}(\text{Classes}))$

Query Result:

class	countries
Iowa	USA
North Carolina	USA
Yamato	Japan

b) **Relational Algebra:**

$\sigma_{launched<1921}(\text{Ships})$

Query Result:

name	class	launched
Haruna	Kongo	1915
Hiei	Kongo	1914
Kirishima	Kongo	1915
Kongo	Kongo	1913
Ramillies	Revenge	1917
Renown	Renown	1916
Repulse	Renown	1916
Resolution	Revenge	1916
Revenge	Revenge	1916
Royal Oak	Revenge	1916
Royal Sovereign	Revenge	1916
Tennessee	Tennessee	1920

c) **Relational Algebra:**

$\sigma_{battle=\text{'Denmark Strait'} \textbf{ AND } result=\text{'sunk'}}(\text{Outcome})$

Query Result:

Ships	battle	result
Bismark	Denmark Strait	sunk
Hood	Denmark Strait	sunk

d) **Relational Algebra:**

$Classes \bowtie_{displacement>35{,}000} Ships$

Query Result:

name	class	launched	type	country	numGuns	bore	displacement
Iowa	Iowa	1943	bb	USA	9	16	46000
Missouri	Iowa	1944	bb	USA	9	16	46000
New Jersey	Iowa	1943	bb	USA	9	16	46000
Wisconsin	Iowa	1944	bb	USA	9	16	46000
Haruna	Kongo	1915	bc	Japan	8	14	32000
Hiei	Kongo	1914	bc	Japan	8	14	32000
Kirishima	Kongo	1915	bc	Japan	8	14	32000
Kongo	Kongo	1913	bc	Japan	8	14	32000
Kongo	Kongo	1913	bc	Japan	8	14	32000
North Carolina	North Carolina	1941	bb	USA	9	16	37000
Washington	North Carolina	1941	bb	USA	9	16	37000
Washington	North Carolina	1941	bb	USA	9	16	37000
Renown	Renown	1916	bc	Gt. Britain	6	15	42000
Repulse	Repulse	1916	bc	Gt. Britain	6	15	42000
Ramillies	Revenge	1917	bb	Gt. Britain	8	15	29000
Resolution	Revenge	1916	bb	Gt. Britain	8	15	29000
Revenge	Revenge	1916	bb	Gt. Britain	8	15	29000
Royal Oak	Revenge	1916	bb	Gt. Britain	8	15	29000
Royal Sovereign	Revenge	1916	bb	Gt. Britain	8	15	29000
California	Tennessee	1921	bb	USA	12	14	32000
Yamato	Yamato	1941	bb	Japan	9	18	65000

Correct Solution:

Relational Algebra:

$Classes \bowtie_{displacement>35,000} Ships$

Query Result:

name	class	launched	type	country	numGuns	bore	displacement
Iowa	Iowa	1943	bb	USA	9	16	46000
Missouri	Iowa	1944	bb	USA	9	16	46000
New Jersey	Iowa	1943	bb	USA	9	16	46000
Wisconsin	Iowa	1944	bb	USA	9	16	46000
Haruna	Kongo	1915	bc	Japan	8	14	32000
Hiei	Kongo	1914	bc	Japan	8	14	32000
Kirishima	Kongo	1915	bc	Japan	8	14	32000
North Carolina	North Carolina	1941	bb	USA	9	16	37000
Washington	North Carolina	1941	bb	USA	9	16	37000
Washington	North Carolina	1941	bb	USA	9	16	37000
Yamato	Yamato	1941	bb	Japan	9	18	65000
Musashi	Yamato	1942	bb	Japan	9	18	65000

e) **Relational Algebra:**

$$\pi_{name,displacement,numGuns}(\text{Classes} \bowtie_{battle=\text{`Guadalcanal}} (\rho_{ship \rightarrow name}(\text{Outcomes} \bowtie \text{Ships})))$$

Query Result:

name	displacement	numGuns
Kirishima	32000	8
Washington	37000	9

f) **Relational Algebra:**

$\pi_{name}(\rho_{ship \rightarrow name}(\text{Outcomes})) \cup \pi_{name}(\text{Ships})$

Query Result:

name
Arizona
Bismark
California
Duke of York
Fuso
Hood
King George V
Kirishima
Prince of Wales
Rodney
Scharnhorst
South Dakota
Tennessee
Washington
West Virginia
Yamashiro
Haruna
Hiei
Iowa
Kongo
Missouri
Musashi
New Jersey
North Carolina
Ramillies
Renown
Repulse
Resolution
Revenge
Royal Oak
Royal Sovereign
Wisconsin
Yamato

g) Omitted for now

h) **Relational Algebra:**

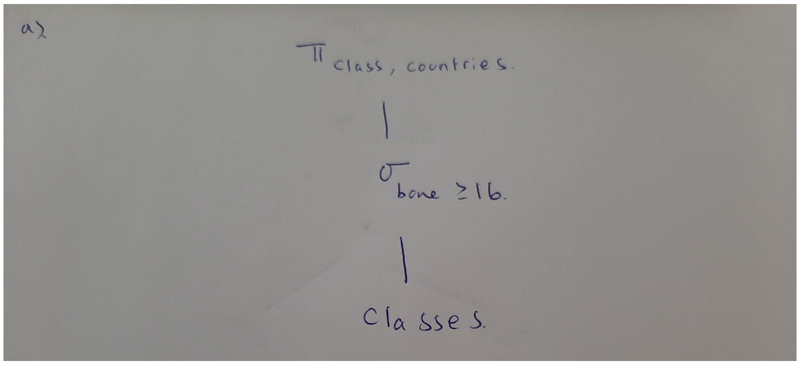
$$\pi_{country}(\sigma_{type=\text{`bb'}}(\text{Classes})) \cap \pi_{country}(\sigma_{type=\text{`bc'}}(\text{Classes}))$$

Query Result:

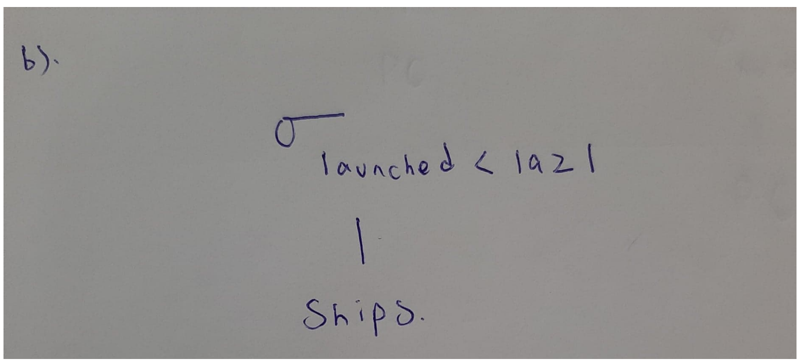
country
Japan
Gt. Britain

i) Omitted for now

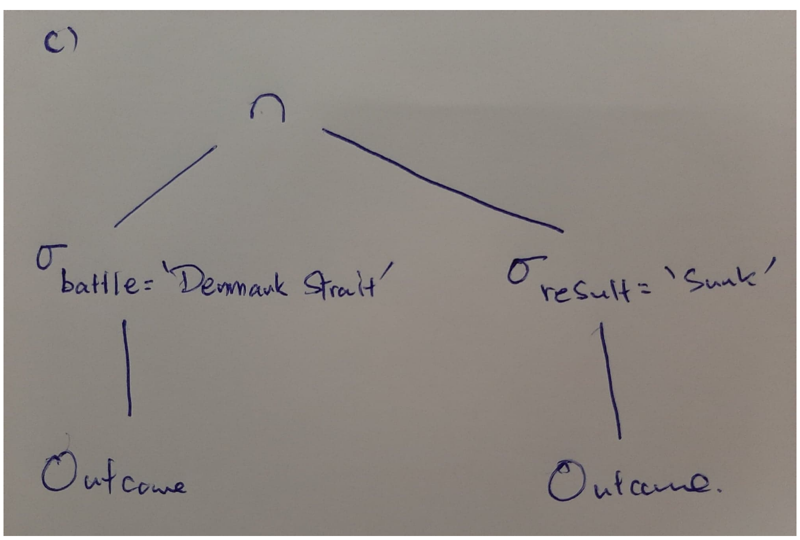
4. a) **Answer:**



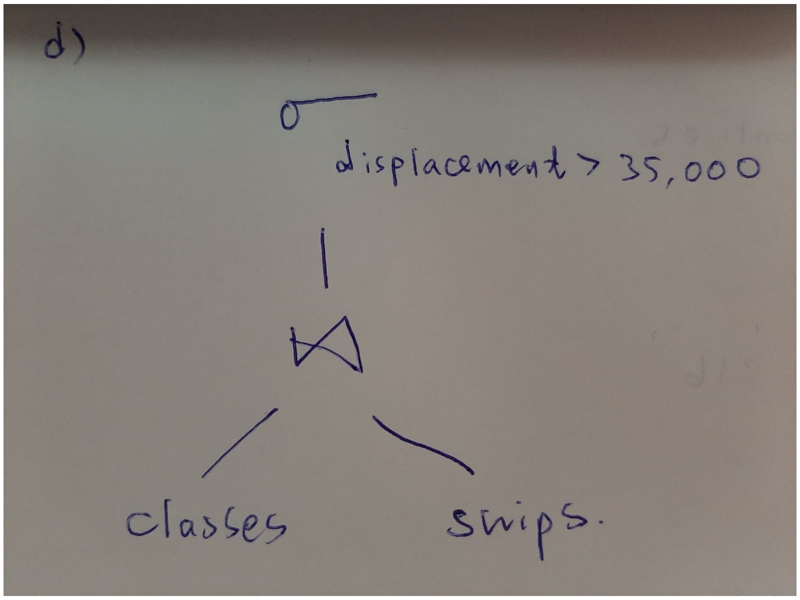
b) **Answer:**



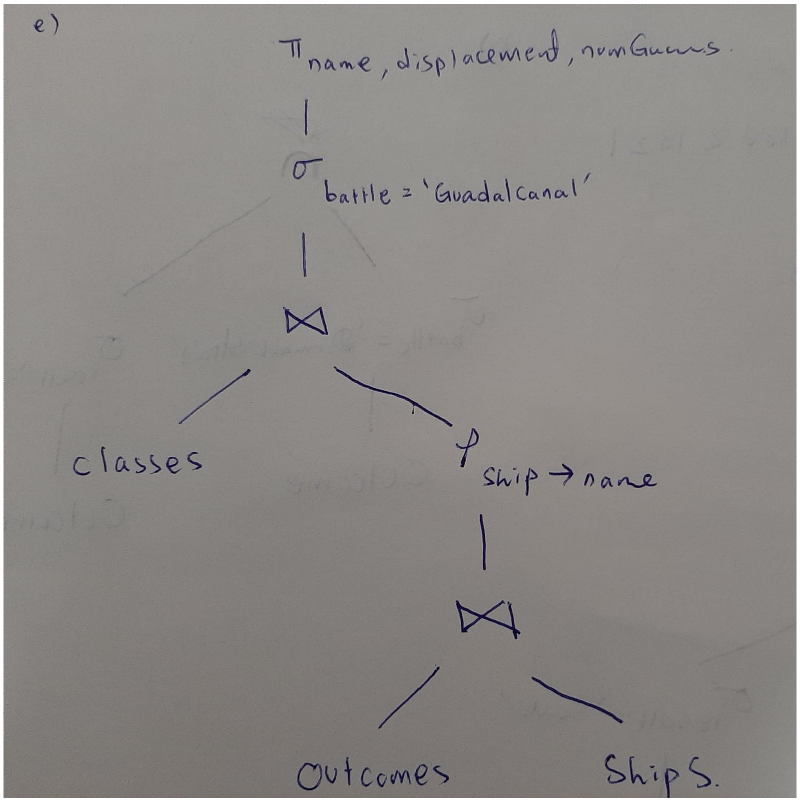
c) **Answer:**



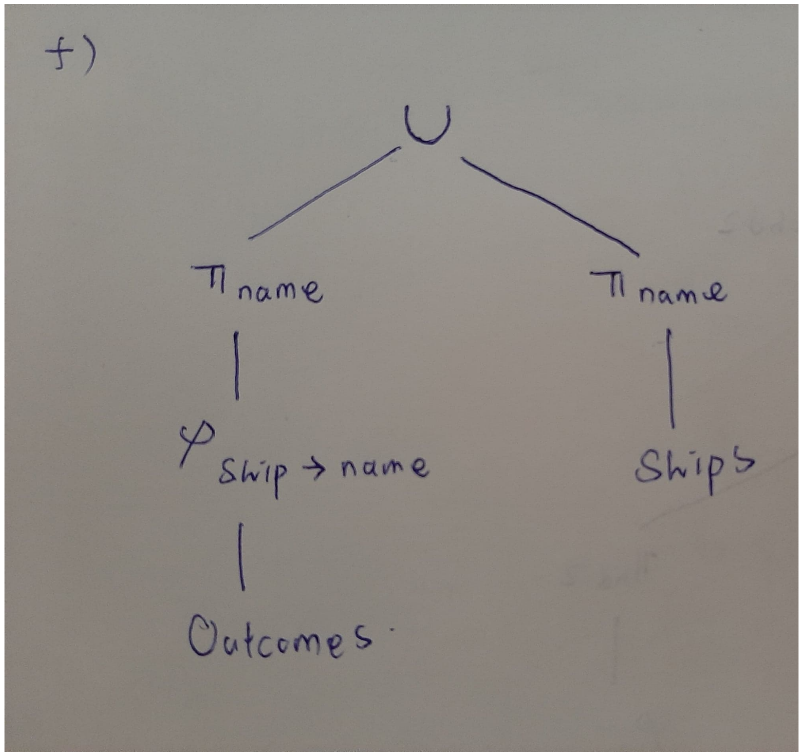
d) **Answer:**



e) **Answer:**

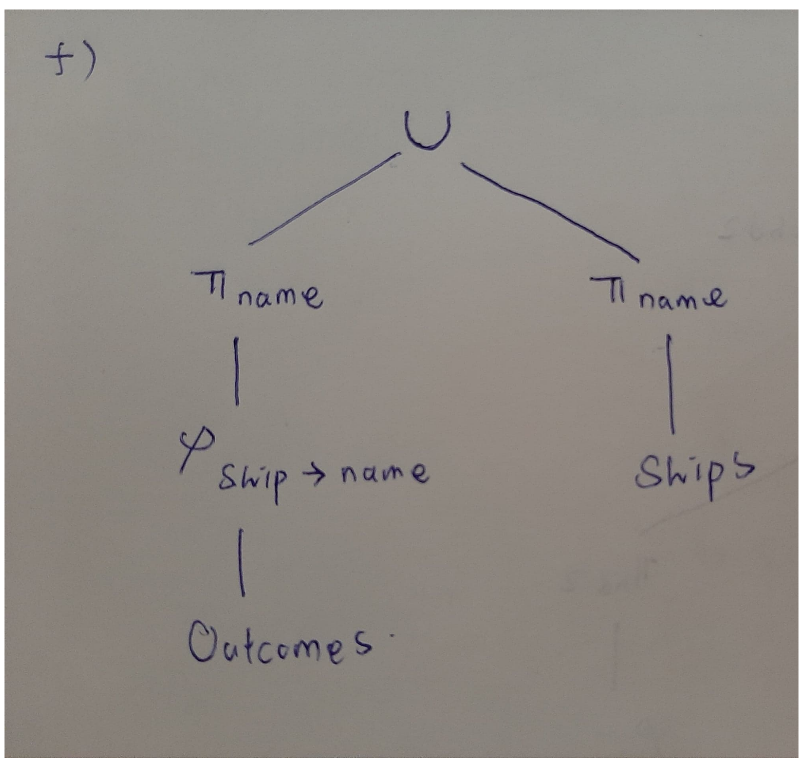


f) **Answer:**



g) Omitted for now

h) **Answer:**



i) Omitted for now

5. One difference exists in the number of attributes.

   The definition of natural join (i.e. $\bowtie$) tells us a copy of duplicate attributes is eliminated.

   Since R and S both have attribute A, the result $R \bowtie S$ would have one attribute of A.

   On the other hand, the definition of theta join tells us $R \bowtie_C S$ is equivalent form of $\sigma_C(R \times S)$.

   Since we know cross product doesn't eliminate duplicate attributes, the result $\sigma_{\text{R.A=S.A}}(R \times S)$ would have 2 attributes of A.

6. **Notes:**

   - **Set:** is an unordered collection of elements without duplicates
     - e.g.

       An example of set

A	B
1	2
3	4
1	2
1	2

   - **Bags:** is unordered collection of elements with duplicates. A bag is also called **Multisets**
     - e.g.

       An example of bag

A	B
1	2
3	4
1	2
1	2