Today

- More relational algebra
  - Division
  - Examples
- SQL
  - Review
  - Examples

Division

- Let R be a relation with fields x, y
- Let S be a relation with field y
- R/S returns one column corresponding to certain x values
- R/S is the set of all x values such that ***for every y value*** in S, there is a tuple (x,y) in R
- Stated another way, for each x value in R, consider the set of all y values in R that correspond to that x value. If this set contains all the values in S, then x is in R/S
- R/S is the largest relation instance Q such that Q x S $\subseteq$ R

Division

R

Supplier_no	Part_no
s1	p1
s1	p2
s1	p3
s1	p4
s2	p1
s2	p2
s3	p2
s4	p2
s4	p4

S1

Part_no
p2

S2

Part_no
p2
p4

S3

Part_no
p1
p2
p4

R/S1

Supplier_no
s1
s2
s3
s4

R/S2

Supplier_no
s1
s4

R/S3

Supplier_no
s1

R / S computes the suppliers who supply all the parts in S

Division

- Generalizing to to a set of attributes
- Let R be a relation with fields $A_1, A_2, \ldots, A_k, A_{k+1}, A_{k+2}, \ldots A_{k+n}$
- Let S be a relation with fields $A_{k+1}, A_{k+2}, \ldots A_{k+n}$
- R/S results in a relation with fields $A_1, A_2, \ldots, A_k$
- R/S is the set of all $A_1, A_2, \ldots, A_k$ such that *for every* $A_{k+1}, A_{k+2}, \ldots A_{k+n}$ in S, there is a tuple in R

sid	sname	rating	age
22	Dustin	7	45.0
29	Brutus	1	33.0
31	Lubber	8	55.5
32	Andy	8	25.5
58	Rusty	10	35.0
64	Horatio	7	35.0
71	Zorba	10	16.0
74	Horatio	9	35.0
85	Art	3	25.5
95	Bob	3	63.5

S3 (Sailors)

sid	bid	day
22	101	10/10/98
22	102	10/10/98
22	103	10/8/98
22	104	10/7/98
31	102	11/10/98
31	103	11/6/98
31	104	11/12/98
64	101	9/5/98
64	102	9/8/98
74	103	9/8/98

R2 (Reserves)

bid	bname	color
101	Interlake	blue
102	Interlake	red
103	Clipper	green
104	Marine	red

B1 (Boats)

Q1: Find the names of sailors who have reserved boat 103

A1: $\pi_{sname}(\sigma_{bid=103}(Reserves) \bowtie Sailors)$

sid	sname	rating	age
22	Dustin	7	45.0
29	Brutus	1	33.0
31	Lubber	8	55.5
32	Andy	8	25.5
58	Rusty	10	35.0
64	Horatio	7	35.0
71	Zorba	10	16.0
74	Horatio	9	35.0
85	Art	3	25.5
95	Bob	3	63.5

S3 (Sailors)

sid	bid	day
22	101	10/10/98
22	102	10/10/98
22	103	10/8/98
22	104	10/7/98
31	102	11/10/98
31	103	11/6/98
31	104	11/12/98
64	101	9/5/98
64	102	9/8/98
74	103	9/8/98

R2 (Reserves)

bid	bname	color
101	Interlake	blue
102	Interlake	red
103	Clipper	green
104	Marine	red

B1 (Boats)

Q2: Find the names of sailors who have reserved a red boat.

A2: $\pi_{sname}(\sigma_{color=red}(Boats) \bowtie Reserves \bowtie Sailors)$

sid	sname	rating	age
22	Dustin	7	45.0
29	Brutus	1	33.0
31	Lubber	8	55.5
32	Andy	8	25.5
58	Rusty	10	35.0
64	Horatio	7	35.0
71	Zorba	10	16.0
74	Horatio	9	35.0
85	Art	3	25.5
95	Bob	3	63.5

S3 (Sailors)

sid	bid	day
22	101	10/10/98
22	102	10/10/98
22	103	10/8/98
22	104	10/7/98
31	102	11/10/98
31	103	11/6/98
31	104	11/12/98
64	101	9/5/98
64	102	9/8/98
74	103	9/8/98

R2 (Reserves)

bid	bname	color
101	Interlake	blue
102	Interlake	red
103	Clipper	green
104	Marine	red

B1 (Boats)

Q3: Find the names of sailors who have reserved a red or a green boat.

A3: $\rho\left(tempboats, (\sigma_{color=red}(Boats) \cup (\sigma_{color=green}(Boats)\right)$

$\pi_{sname}(tempboats) \quad \bowtie Reserves \bowtie$Sailors)

sid	sname	rating	age
22	Dustin	7	45.0
29	Brutus	1	33.0
31	Lubber	8	55.5
32	Andy	8	25.5
58	Rusty	10	35.0
64	Horatio	7	35.0
71	Zorba	10	16.0
74	Horatio	9	35.0
85	Art	3	25.5
95	Bob	3	63.5

S3 (Sailors)

sid	bid	day
22	101	10/10/98
22	102	10/10/98
22	103	10/8/98
22	104	10/7/98
31	102	11/10/98
31	103	11/6/98
31	104	11/12/98
64	101	9/5/98
64	102	9/8/98
74	103	9/8/98

R2 (Reserves)

bid	bname	color
101	Interlake	blue
102	Interlake	red
103	Clipper	green
104	Marine	red

B1 (Boats)

Q4: Find the colors of boats reserved by Lubber.

A4: $\pi_{color}(\sigma_{sname=Lubber}(Sailors) \bowtie Reserves \bowtie Boats)$

sid	sname	rating	age
22	Dustin	7	45.0
29	Brutus	1	33.0
31	Lubber	8	55.5
32	Andy	8	25.5
58	Rusty	10	35.0
64	Horatio	7	35.0
71	Zorba	10	16.0
74	Horatio	9	35.0
85	Art	3	25.5
95	Bob	3	63.5

S3 (Sailors)

sid	bid	day
22	101	10/10/98
22	102	10/10/98
22	103	10/8/98
22	104	10/7/98
31	102	11/10/98
31	103	11/6/98
31	104	11/12/98
64	101	9/5/98
64	102	9/8/98
74	103	9/8/98

R2 (Reserves)

bid	bname	color
101	Interlake	blue
102	Interlake	red
103	Clipper	green
104	Marine	red

B1 (Boats)

Q5: Find the names of sailors who have reserved a red and a green boat.

A5: $\rho(tempred, \pi_{sid}((\sigma_{color=red}(Boats)) \bowtie Reserves))$

$\rho(tempgreen, \pi_{sid}((\sigma_{color=green}(Boats)) \bowtie Reserves))$

$\pi_{sname}(tempred \cap tempgreen)$ ⋈Sailors)

sid	sname	rating	age
22	Dustin	7	45.0
29	Brutus	1	33.0
31	Lubber	8	55.5
32	Andy	8	25.5
58	Rusty	10	35.0
64	Horatio	7	35.0
71	Zorba	10	16.0
74	Horatio	9	35.0
85	Art	3	25.5
95	Bob	3	63.5

S3 (Sailors)

sid	bid	day
22	101	10/10/98
22	102	10/10/98
22	103	10/8/98
22	104	10/7/98
31	102	11/10/98
31	103	11/6/98
31	104	11/12/98
64	101	9/5/98
64	102	9/8/98
74	103	9/8/98

R2 (Reserves)

bid	bname	color
101	Interlake	blue
102	Interlake	red
103	Clipper	green
104	Marine	red

B1 (Boats)

Q: Find the names of sailors who have reserved at least one boat.

A: $\pi_{sname}(Sailors \bowtie$Reserves)

sid	sname	rating	age
22	Dustin	7	45.0
29	Brutus	1	33.0
31	Lubber	8	55.5
32	Andy	8	25.5
58	Rusty	10	35.0
64	Horatio	7	35.0
71	Zorba	10	16.0
74	Horatio	9	35.0
85	Art	3	25.5
95	Bob	3	63.5

S3 (Sailors)

sid	bid	day
22	101	10/10/98
22	102	10/10/98
22	103	10/8/98
22	104	10/7/98
31	102	11/10/98
31	103	11/6/98
31	104	11/12/98
64	101	9/5/98
64	102	9/8/98
74	103	9/8/98

R2 (Reserves)

bid	bname	color
101	Interlake	blue
102	Interlake	red
103	Clipper	green
104	Marine	red

B1 (Boats)

Q: Find the names of sailors who have reserved at least two boats.

A: $\rho\left(reservations, \pi_{sid,name,bid}(Sailors \bowtie Reserves)\right)$

$\rho(reservationPairs(1 \to sid1, 2 \to sname1, 3 \to bid1, 4 \to sid2,$

$5 \to sname2, 6 \to bid2), reservations \times reservations)$

$\pi_{sname1}(\sigma_{sid1=sid2 \wedge (bid1 \neq bid2)}(reservationPairs))$

sid	sname	rating	age
22	Dustin	7	45.0
29	Brutus	1	33.0
31	Lubber	8	55.5
32	Andy	8	25.5
58	Rusty	10	35.0
64	Horatio	7	35.0
71	Zorba	10	16.0
74	Horatio	9	35.0
85	Art	3	25.5
95	Bob	3	63.5

S3 (Sailors)

sid	bid	day
22	101	10/10/98
22	102	10/10/98
22	103	10/8/98
22	104	10/7/98
31	102	11/10/98
31	103	11/6/98
31	104	11/12/98
64	101	9/5/98
64	102	9/8/98
74	103	9/8/98

R2 (Reserves)

bid	bname	color
101	Interlake	blue
102	Interlake	red
103	Clipper	green
104	Marine	red

B1 (Boats)

Q: Find the sids of sailors with age over 20 who have not reserved a red boat

A: $\pi_{sid}(\sigma_{age>20}Sailors) - \pi_{sid}(\sigma_{color=red}(Boats) \bowtie Reserves \bowtie Sailors)$

sid	sname	rating	age
22	Dustin	7	45.0
29	Brutus	1	33.0
31	Lubber	8	55.5
32	Andy	8	25.5
58	Rusty	10	35.0
64	Horatio	7	35.0
71	Zorba	10	16.0
74	Horatio	9	35.0
85	Art	3	25.5
95	Bob	3	63.5

S3 (Sailors)

sid	bid	day
22	101	10/10/98
22	102	10/10/98
22	103	10/8/98
22	104	10/7/98
31	102	11/10/98
31	103	11/6/98
31	104	11/12/98
64	101	9/5/98
64	102	9/8/98
74	103	9/8/98

R2 (Reserves)

bid	bname	color
101	Interlake	blue
102	Interlake	red
103	Clipper	green
104	Marine	red

B1 (Boats)

Q: Find the names of sailors who have reserved all boats

A: $\rho(tempsids, \pi_{sid,bid} Reserves / \pi_{bid} Boats)$

$\pi_{sname}(tempsids \bowtie Sailors)$

sid	sname	rating	age
22	Dustin	7	45.0
29	Brutus	1	33.0
31	Lubber	8	55.5
32	Andy	8	25.5
58	Rusty	10	35.0
64	Horatio	7	35.0
71	Zorba	10	16.0
74	Horatio	9	35.0
85	Art	3	25.5
95	Bob	3	63.5

S3 (Sailors)

sid	bid	day
22	101	10/10/98
22	102	10/10/98
22	103	10/8/98
22	104	10/7/98
31	102	11/10/98
31	103	11/6/98
31	104	11/12/98
64	101	9/5/98
64	102	9/8/98
74	103	9/8/98

R2 (Reserves)

bid	bname	color
101	Interlake	blue
102	Interlake	red
103	Clipper	green
104	Marine	red

B1 (Boats)

Q: Find the names of sailors who have reserved all boats called Interlake

sid	sname	rating	age
22	Dustin	7	45.0
29	Brutus	1	33.0
31	Lubber	8	55.5
32	Andy	8	25.5
58	Rusty	10	35.0
64	Horatio	7	35.0
71	Zorba	10	16.0
74	Horatio	9	35.0
85	Art	3	25.5
95	Bob	3	63.5

S3 (Sailors)

sid	bid	day
22	101	10/10/98
22	102	10/10/98
22	103	10/8/98
22	104	10/7/98
31	102	11/10/98
31	103	11/6/98
31	104	11/12/98
64	101	9/5/98
64	102	9/8/98
74	103	9/8/98

R2 (Reserves)

bid	bname	color
101	Interlake	blue
102	Interlake	red
103	Clipper	green
104	Marine	red

B1 (Boats)

Q: Find the names of sailors who have reserved all boats called Interlake

A: $\rho\left(tempsids, \pi_{sid,bid} Reserves / \pi_{bid}\left(\sigma_{bname=\text{“}Interlake\text{”}}(Boats)\right)\right)$

$\pi_{sname}(tempsids \bowtie Sailors)$

SQL

- Basic form:

```
SELECT [DISTINCT] <select-list>
FROM <from-list>
WHERE <qualification>
```

- from-list: names of tables to query from. A table name can be followed by a range variable (“as <alias>”)
- select-list: list of column names or expressions involving common names of tables named in the from-list. Can be prefixed by a range variable (alias.<column_name>)
- qualification: Boolean combination of conditions in the form:
  - “expression op expression”
    - Op is one of the comparison operators {<, <=, =, <>, >=, >}
    - Expression is a column name, constant, or an expression
- DISTINCT (optional) makes results unique (a set) -> results to SQL queries contain duplicates by default!

sid	sname	rating	age
22	Dustin	7	45.0
29	Brutus	1	33.0
31	Lubber	8	55.5
32	Andy	8	25.5
58	Rusty	10	35.0
64	Horatio	7	35.0
71	Zorba	10	16.0
74	Horatio	9	35.0
85	Art	3	25.5
95	Bob	3	63.5

S3 (Sailors)

sid	bid	day
22	101	10/10/98
22	102	10/10/98
22	103	10/8/98
22	104	10/7/98
31	102	11/10/98
31	103	11/6/98
31	104	11/12/98
64	101	9/5/98
64	102	9/8/98
74	103	9/8/98

R2 (Reserves)

bid	bname	color
101	Interlake	blue
102	Interlake	red
103	Clipper	green
104	Marine	red

B1 (Boats)

Q: Find the names and ages of all sailors.

sid	sname	rating	age
22	Dustin	7	45.0
29	Brutus	1	33.0
31	Lubber	8	55.5
32	Andy	8	25.5
58	Rusty	10	35.0
64	Horatio	7	35.0
71	Zorba	10	16.0
74	Horatio	9	35.0
85	Art	3	25.5
95	Bob	3	63.5

S3 (Sailors)

sid	bid	day
22	101	10/10/98
22	102	10/10/98
22	103	10/8/98
22	104	10/7/98
31	102	11/10/98
31	103	11/6/98
31	104	11/12/98
64	101	9/5/98
64	102	9/8/98
74	103	9/8/98

R2 (Reserves)

bid	bname	color
101	Interlake	blue
102	Interlake	red
103	Clipper	green
104	Marine	red

B1 (Boats)

Q: Find the names and ages of all sailors.

```
SELECT DISTINCT S.name, S.age
FROM Sailors S
```

What if we don't include DISTINCT?

sid	sname	rating	age
22	Dustin	7	45.0
29	Brutus	1	33.0
31	Lubber	8	55.5
32	Andy	8	25.5
58	Rusty	10	35.0
64	Horatio	7	35.0
71	Zorba	10	16.0
74	Horatio	9	35.0
85	Art	3	25.5
95	Bob	3	63.5

S3 (Sailors)

sid	bid	day
22	101	10/10/98
22	102	10/10/98
22	103	10/8/98
22	104	10/7/98
31	102	11/10/98
31	103	11/6/98
31	104	11/12/98
64	101	9/5/98
64	102	9/8/98
74	103	9/8/98

R2 (Reserves)

bid	bname	color
101	Interlake	blue
102	Interlake	red
103	Clipper	green
104	Marine	red

B1 (Boats)

Q: Find all sailors with a rating above 7.

sid	sname	rating	age
22	Dustin	7	45.0
29	Brutus	1	33.0
31	Lubber	8	55.5
32	Andy	8	25.5
58	Rusty	10	35.0
64	Horatio	7	35.0
71	Zorba	10	16.0
74	Horatio	9	35.0
85	Art	3	25.5
95	Bob	3	63.5

S3 (Sailors)

sid	bid	day
22	101	10/10/98
22	102	10/10/98
22	103	10/8/98
22	104	10/7/98
31	102	11/10/98
31	103	11/6/98
31	104	11/12/98
64	101	9/5/98
64	102	9/8/98
74	103	9/8/98

R2 (Reserves)

bid	bname	color
101	Interlake	blue
102	Interlake	red
103	Clipper	green
104	Marine	red

B1 (Boats)

Q: Find all sailors with a rating above 7.

```
SELECT S.sid, S.sname, S.rating, S.age
FROM Sailors as S
WHERE S.rating > 7
```

Can also use SELECT *

SQL – Set operations

- ***UNION***: Set union
- ***INTERSECT***: Set intersection
- ***EXCEPT***: Set difference
- In contrast to other SQL queries, these operations remove duplicates by default. To include duplicates, need to include ALL keyword:
  - UNION ALL
  - INTERSECT ALL
  - EXCEPT ALL

sid	sname	rating	age
22	Dustin	7	45.0
29	Brutus	1	33.0
31	Lubber	8	55.5
32	Andy	8	25.5
58	Rusty	10	35.0
64	Horatio	7	35.0
71	Zorba	10	16.0
74	Horatio	9	35.0
85	Art	3	25.5
95	Bob	3	63.5

S3 (Sailors)

sid	bid	day
22	101	10/10/98
22	102	10/10/98
22	103	10/8/98
22	104	10/7/98
31	102	11/10/98
31	103	11/6/98
31	104	11/12/98
64	101	9/5/98
64	102	9/8/98
74	103	9/8/98

R2 (Reserves)

bid	bname	color
101	Interlake	blue
102	Interlake	red
103	Clipper	green
104	Marine	red

B1 (Boats)

Q3: Find the names of sailors who have reserved a red or a green boat.

```
SELECT S.sname
FROM   Sailors S, Reserves R, Boats B
WHERE  S.sid = R.sid AND R.bid = B.bid
       AND (B.color = 'red' OR B.color = 'green')
```

sid	sname	rating	age

S3 (Sailors)

sid	bid	day

R2 (Reserves)

bid	bname	color

B1 (Boats)

Q5: Find the names of sailors who have reserved a red and a green boat.

```
SELECT S.sname
FROM Sailors S, Reserves R, Boats B
WHERE S.sicl = R.sid AND R.bid = B.bid AND B.color = 'red'
UNION
SELECT S2.sname
FROM Sailors S2, Boats B2, Reserves H2
WHERE S2.sid = H2.sid AND R2.bid = B2.bicl AND B2.color =
'green'
```

Nested Queries

- A nested query is a query with another query embedded in it
- Also known as a “subquery”
- Useful to break complicated queries into smaller queries

sid	sname	rating	age

S3 (Sailors)

sid	bid	day

R2 (Reserves)

bid	bname	color

B1 (Boats)

Q1: Find the names of sailors who have reserved boat 103

```
SELECT S.name
FROM Sailors S
WHERE S.sid IN (SELECT R.sid
                FROM Reserves R
                WHERE R.bid = 103)
```

sid	sname	rating	age

S3 (Sailors)

sid	bid	day

R2 (Reserves)

bid	bname	color

B1 (Boats)

Q2: Find the names of sailors who have reserved a red boat

```
SELECT S.name
FROM Sailors S
WHERE S.sid IN (SELECT R.sid
                FROM Reserves R
                WHERE R.bid IN (SELECT B.bid
                                FROM Boats B
                                WHERE B.color = ‘red’))
```

Aggregate Operators

- COUNT ([DISTINCT A): The number of [unique] values in column A
- SUM ([DISTINCT A]): The sum of all [unique] values in column A
- AVG ([DISTINCT A]): The average of all [unique] values in column A
- MAX (A): The maximum of all values in column A
- MIN (A): The minimum of all values in column A

sid	sname	rating	age

S3 (Sailors)

sid	bid	day

R2 (Reserves)

bid	bname	color

B1 (Boats)

Find the average age of all sailors.

```
SELECT AVG(S.age) FROM SAILORS S
```

Find the average age of sailors with a rating of 10

```
SELECT AVG(S.age) FROM SAILORS S WHERE S.rating = 10
```

Find the name and age of the oldest sailor

```
SELECT S.sname, S.age
FROM Sailors S
WHERE S.age = (SELECT MAX(S2.age) FROM SAILOR s2)
```

Count the number of sailors

```
SELECT COUNT(*) FROM Sailors S
```

Count the number of different sailor names.

```
SELECT COUNT(DISTINCT S.sname) FROM Sailors S
```

GROUP BY and HAVING

- SELECT [DISTINCT] <select-list>
- FROM <from-list>
- WHERE <qualification>
- GROUP BY <grouping-list>
- HAVING <group-qualification>

- <grouping-list>: Used when we want to apply aggregate operations to each member of a group of rows in a relation. Specifies the list of fields we want to group by.
- <group-qualification>: condition on the grouping

sid	sname	rating	age

S3 (Sailors)

sid	bid	day

R2 (Reserves)

bid	bname	color

B1 (Boats)

Find the age of the youngest sailor for each rating level.

```
SELECT min(S.age)
FROM Sailors S
WHERE S.rating = i
```

Need to specifiy i = 1,2, 3, … 10. Another way:

```
SELECT S.rating, MIN(S.age) as min_age
FROM Sailors S
GROUP BY S.rating
```

If we only want to include rows where the minimum age is > 40

```
SELECT S.rating, MIN(S.age) as min_age
FROM Sailors S
GROUP BY S.rating
HAVING min_age > 40
```