Aggregations and Groupings

CS 327E
October 9, 2017

Announcements:

- Keep track of problems encountered during Labs
- Next week: Lab 3
- Before next class: setup QuickSight using guide

1)Which of the following is an **aggregate** function?

A) COUNT
B) TRIM
C) DISTINCT
D) All of the above

2) Given the Employee_Bonus table shown, we want to find the single highest bonus amount paid out to an employee. Will the following query compute the right answer?

```
SELECT MAX(SUM(amount))
FROM Employee_Bonus;
```

A) TRUE
B) FALSE

Employee_Bonus

empid	quarter	year	amount
A	3	2017	1000
B	3	2017	2000
A	2	2017	500
C	2	2017	5000
B	2	2017	2000
C	1	2017	800

3) Given the Employee_Bonus table shown, we want to add up all bonuses paid out to an employee and return this total grouped by employee. Will the following query compute the right answer?

```
SELECT empid, SUM(amount)
FROM Employee_Bonus
GROUP BY empid;
```

A) TRUE
B) FALSE

Employee_Bonus

empid	quarter	year	amount
A	3	2017	1000
B	3	2017	2000
A	2	2017	500
C	2	2017	5000
B	2	2017	2000
C	1	2017	800

4) What columns would make a good primary key for the Employee_Bonus table?

A) empid, amount
B) empid, quarter
C) empid, quarter, year
D) empid, year

Employee_Bonus

empid	quarter	year	amount
A	3	2017	1000
B	3	2017	2000
A	2	2017	500
C	2	2017	5000
B	2	2017	2000
C	1	2017	800

5) All aggregate functions except _____ ignore null values in their input set

A) SUM(attribute)
B) COUNT(attribute)
C) AVG(attribute)
D) COUNT(*)

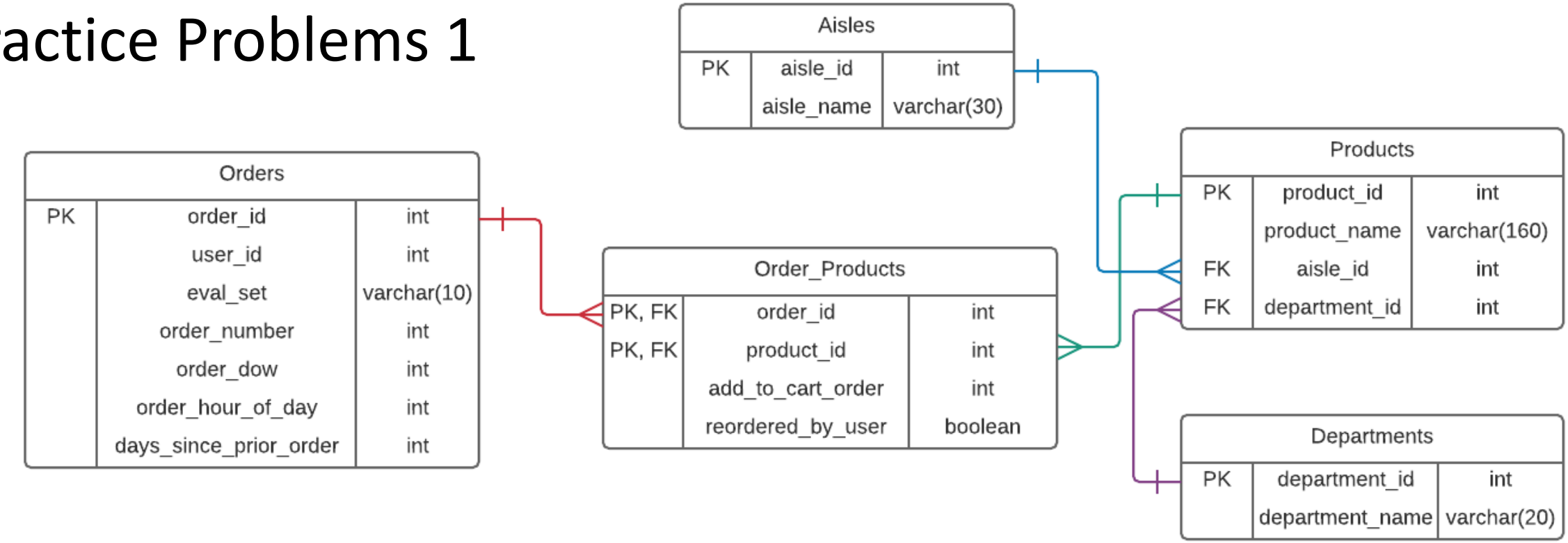
Aggregate Functions

- MIN
- MAX
- SUM
- AVG
- COUNT

Employee

empid	firstname	lastname	salary	depid
1	Michael	Dell	100	5
2	Betty	Jennings	200	
3	Bill	Gates	0	5
4	Kay	McNulty	300	8
5	Jim	Gray	500	6
6	Gordon	Moore	400	6

Practice Problems 1

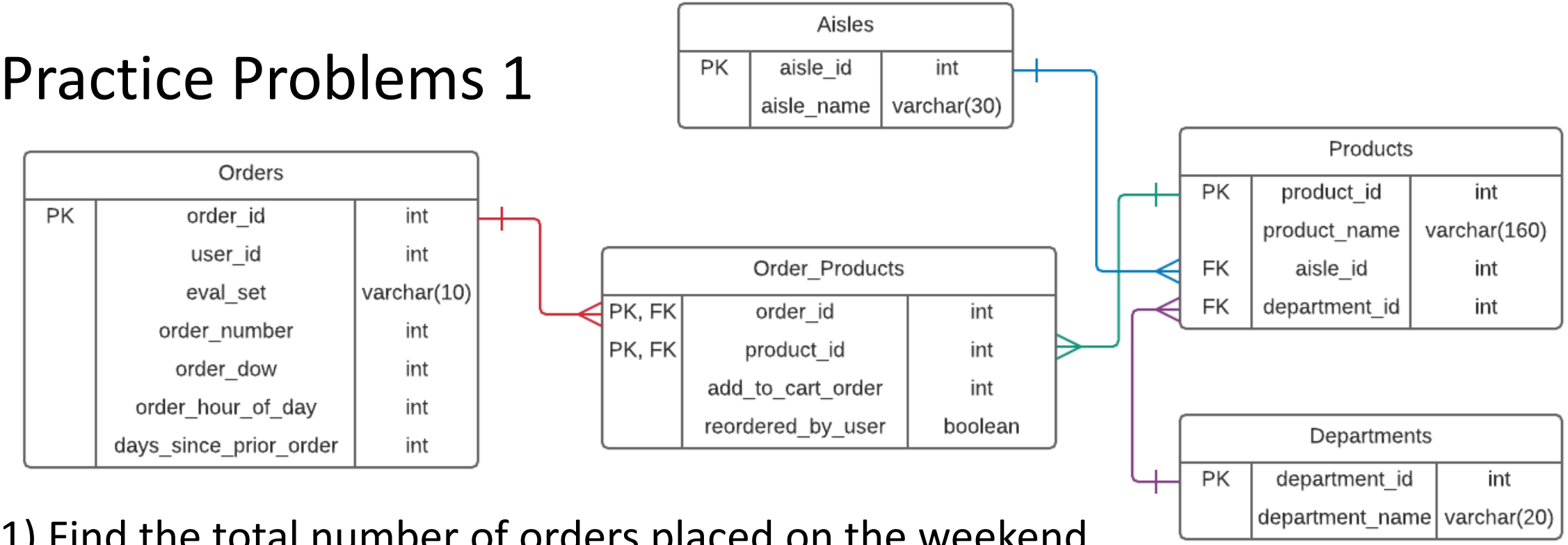


1) Find the total number of orders placed on the weekend

2) Find the highest number of products in an order. Hint: use add_to_cart_order

3) Find the average number of days between orders

Practice Problems 1



1) Find the total number of orders placed on the weekend
2) Find the highest number of products in an order
3) Find the average number of days between orders

Which aggregates did you use to compute these answers?

A) count(*), max, avg B) sum, max, avg C) count(*), sum, avg

Aggregates with Groupings

- MIN
- MAX
- SUM
- AVG
- COUNT

Employee

empid	firstname	lastname	salary	depid
1	Michael	Dell	100	5
2	Betty	Jennings	200	
3	Bill	Gates	0	5
4	Kay	McNulty	300	8
5	Jim	Gray	500	6
6	Gordon	Moore	400	6

```
dev=> select depid, count(*) from Employee group by depid;
 depid | count
-------+-------
       |     1
     8 |     1
     5 |     2
     6 |     2
(4 rows)
```

Aggregates with Groupings

- MIN
- MAX
- SUM
- AVG
- COUNT

Employee

empid	firstname	lastname	salary	depid
1	Michael	Dell	100	5
2	Betty	Jennings	200	
3	Bill	Gates	0	5
4	Kay	McNulty	300	8
5	Jim	Gray	500	6
6	Gordon	Moore	400	6

```
dev=> select depid, sum(salary) from Employee group by depid;
 depid | sum
-------+-----
       | 200
     8 | 300
     5 | 100
     6 | 900
(4 rows)
```

Aggregates with Groupings

- MIN
- MAX
- SUM
- AVG
- COUNT

Employee

empid	firstname	lastname	salary	depid
1	Michael	Dell	100	5
2	Betty	Jennings	200	
3	Bill	Gates	0	5
4	Kay	McNulty	300	8
5	Jim	Gray	500	6
6	Gordon	Moore	400	6

```
dev=> select depname, e.depid, sum(salary)
dev-> from Employee e join Department d on e.depid = d.depid
dev-> group by depname, e.depid;
   depname    | depid | sum
--------------+-------+-----
 Engineering |     8 | 300
 Research    |     6 | 900
 Executive   |     5 | 100
(3 rows)
```

Department

depid	depname
5	Executive
6	Research
7	Sales
8	Engineering

Aggregates with Groupings

- MIN
- MAX
- SUM
- AVG
- COUNT

Employee

empid	firstname	lastname	salary	depid
1	Michael	Dell	100	5
2	Betty	Jennings	200	
3	Bill	Gates	0	5
4	Kay	McNulty	300	8
5	Jim	Gray	500	6
6	Gordon	Moore	400	6

```
dev=> select depname, e.depid, sum(salary)
dev-> from Employee e left outer join Department d on e.depid = d.depid
dev-> group by depname, e.depid;
   depname    | depid | sum
--------------+-------+-----
              |       | 200
 Engineering  |     8 | 300
 Research     |     6 | 900
 Executive    |     5 | 100
(4 rows)
```

Department

depid	depname
5	Executive
6	Research
7	Sales
8	Engineering

Aggregates with Groupings

- MIN
- MAX
- SUM
- AVG
- COUNT

Employee

empid	firstname	lastname	salary	depid
1	Michael	Dell	100	5
2	Betty	Jennings	200	
3	Bill	Gates	0	5
4	Kay	McNulty	300	8
5	Jim	Gray	500	6
6	Gordon	Moore	400	6

```
dev=> select depname, d.depid, sum(salary)
dev-> from Employee e right outer join Department d on e.depid = d.depid
dev-> group by depname, d.depid;
   depname   | depid | sum
-------------+-------+-----
 Sales       |     7 |
 Engineering |     8 | 300
 Executive   |     5 | 100
 Research    |     6 | 900
(4 rows)
```

Department

depid	depname
5	Executive
6	Research
7	Sales
8	Engineering

Aggregates with Groupings

- MIN
- MAX
- SUM
- AVG
- COUNT

Employee

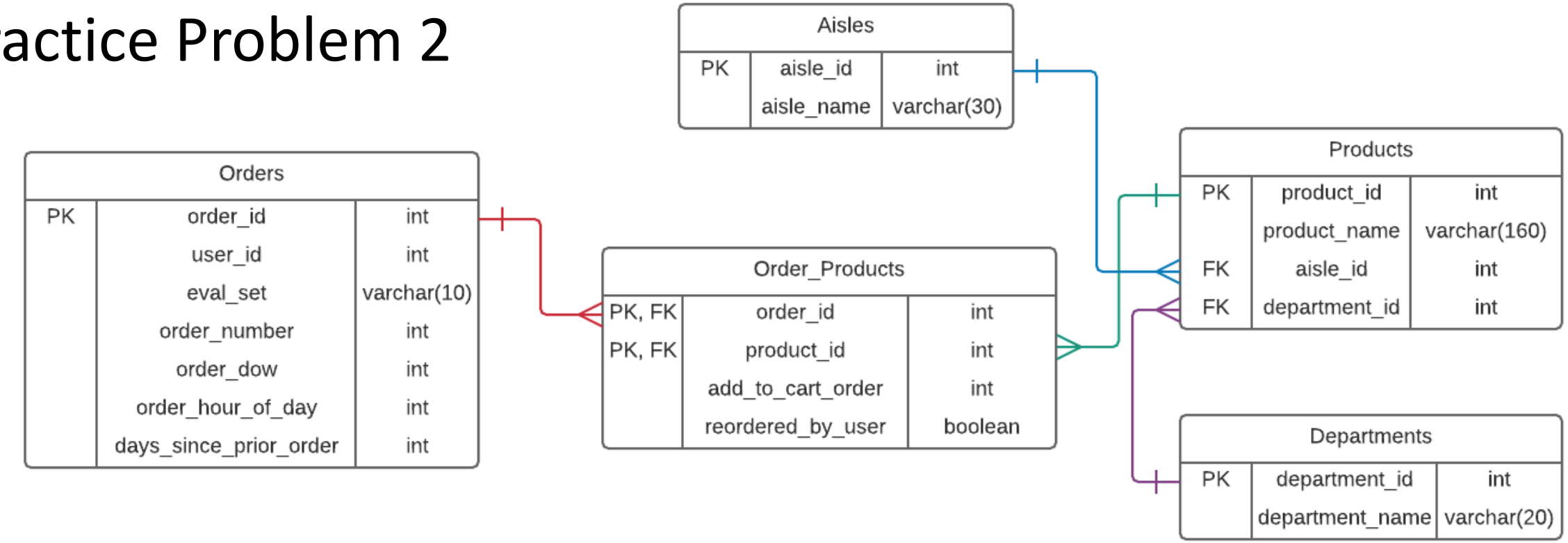
empid	firstname	lastname	salary	depid
1	Michael	Dell	100	5
2	Betty	Jennings	200	
3	Bill	Gates	0	5
4	Kay	McNulty	300	8
5	Jim	Gray	500	6
6	Gordon	Moore	400	6

```
dev=> select depname, d.depid, sum(salary)
dev-> from Employee e full outer join Department d on e.depid = d.depid
dev-> group by depname, d.depid;
   depname    | depid | sum
--------------+-------+-----
              |       | 200
 Sales        |     7 |
 Engineering  |     8 | 300
 Executive    |     5 | 100
 Research     |     6 | 900
(5 rows)
```

Department

depid	depname
5	Executive
6	Research
7	Sales
8	Engineering

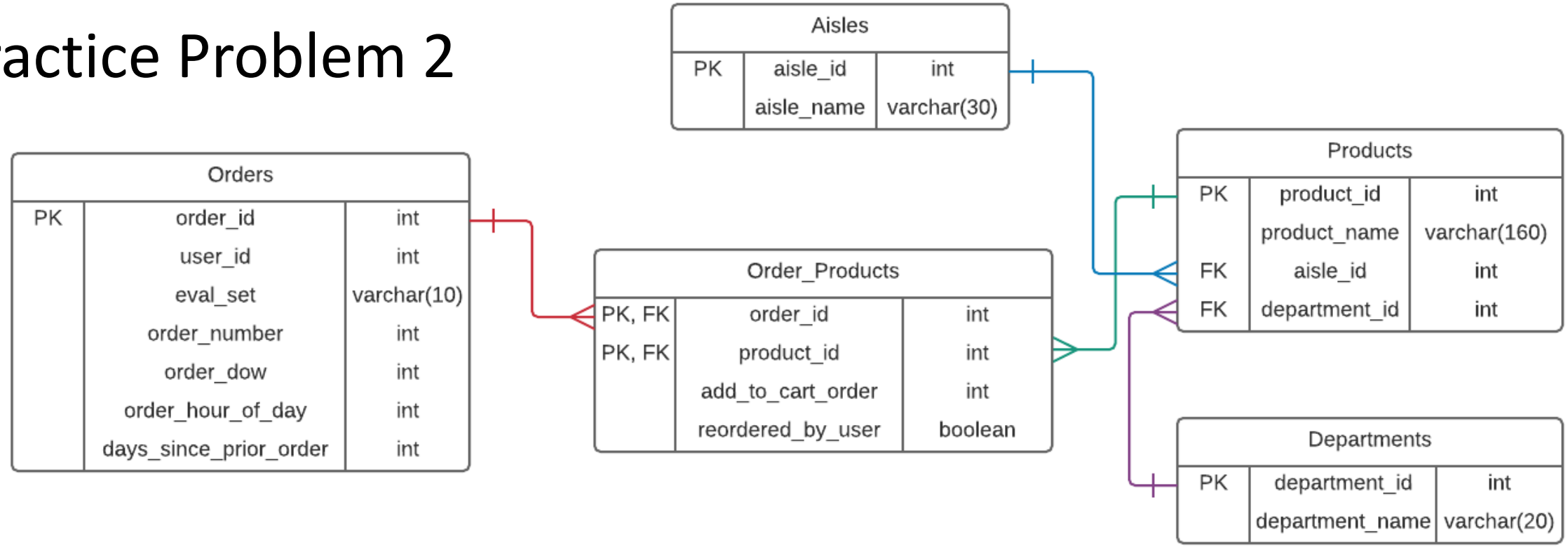
Practice Problem 2



Find the 10 **most** popular products (popular = frequently ordered)

Display the product name and number of times it was ordered

Practice Problem 2



Find the 10 **most** popular products (popular = frequently ordered)

Does the query require a group-by? If so, what is the grouping column?

A) order_id B) product_id C) product_name D) N/A

Semantics of COUNT

```
dev=> select count(*) from Employee;
 count
-------
     6
(1 row)
```

```
dev=> select count(depid) from Employee;
 count
-------
     5
(1 row)
```

```
dev=> select count(distinct depid) from Employee;
 count
-------
     3
(1 row)
```

Employee

empid	firstname	lastname	salary	depid
1	Michael	Dell	100	5
2	Betty	Jennings	200	
3	Bill	Gates	0	5
4	Kay	McNulty	300	8
5	Jim	Gray	500	6
6	Gordon	Moore	400	6

COUNT(*) is Special

- MIN
- MAX
- SUM
- AVG
- COUNT

Employee

empid	firstname	lastname	salary	depid
1	Michael	Dell	100	5
2	Betty	Jennings	200	
3	Bill	Gates	0	5
4	Kay	McNulty	300	8
5	Jim	Gray	500	6
6	Gordon	Moore	400	6

```
dev=> select depname, d.depid, count(*)
dev-> from Employee e right outer join Department d on e.depid = d.depid
dev-> group by depname, d.depid;
   depname    | depid | count
--------------+-------+-------
 Sales        |     7 |     1
 Engineering  |     8 |     1
 Executive    |     5 |     2
 Research     |     6 |     2
(4 rows)
```

Department

depid	depname
5	Executive
6	Research
7	Sales
8	Engineering

COUNT(*) is Special

- MIN
- MAX
- SUM
- AVG
- COUNT

Employee

empid	firstname	lastname	salary	depid
1	Michael	Dell	100	5
2	Betty	Jennings	200	
3	Bill	Gates	0	5
4	Kay	McNulty	300	8
5	Jim	Gray	500	6
6	Gordon	Moore	400	6

```
dev=> select depname, d.depid, count(e.empid)
dev-> from Employee e right outer join Department d on e.depid = d.depid
dev-> group by depname, d.depid;
   depname   | depid | count
-------------+-------+-------
 Sales       |     7 |     0
 Engineering |     8 |     1
 Executive   |     5 |     2
 Research    |     6 |     2
(4 rows)
```

Department

depid	depname
5	Executive
6	Research
7	Sales
8	Engineering

COUNT(*) is Special

- MIN
- MAX
- SUM
- AVG
- COUNT

Employee

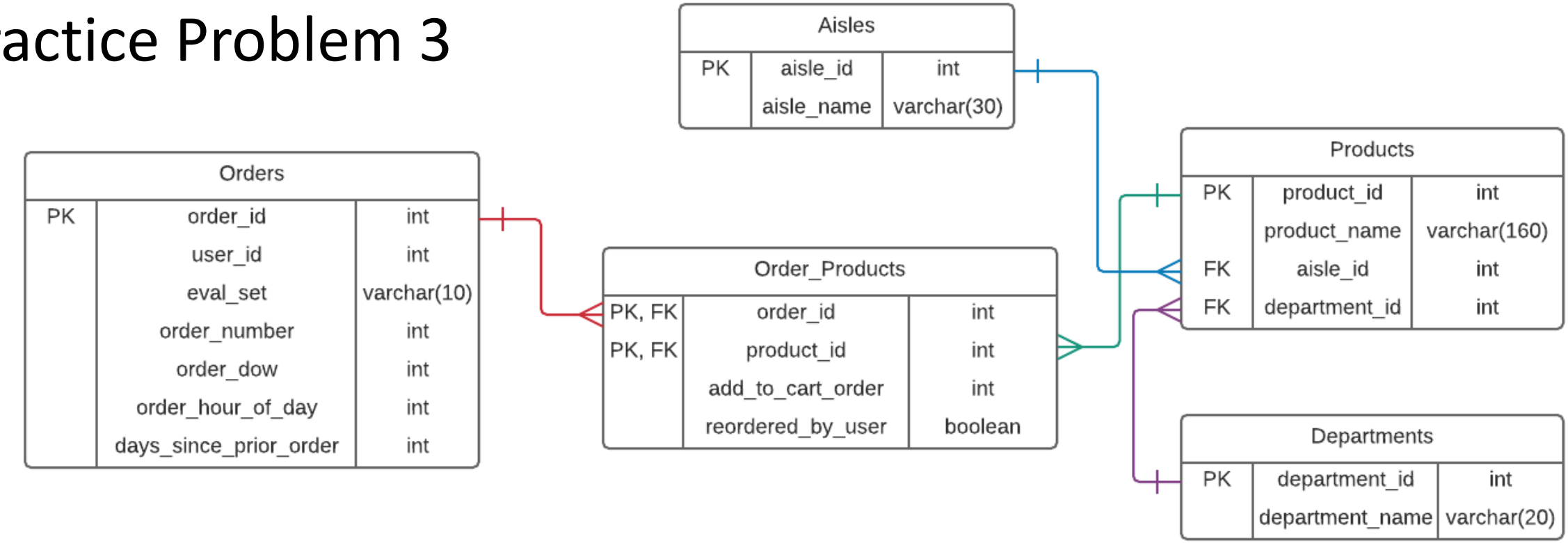
empid	firstname	lastname	salary	depid
1	Michael	Dell	100	5
2	Betty	Jennings	200	
3	Bill	Gates	0	5
4	Kay	McNulty	300	8
5	Jim	Gray	500	6
6	Gordon	Moore	400	6

```
dev=> select depname, d.depid, count(e.empid)
dev-> from Employee e full outer join Department d on e.depid = d.depid
dev-> group by depname, d.depid;
   depname    | depid | count
--------------+-------+-------
              |       |     1
 Sales        |     7 |     0
 Engineering  |     8 |     1
 Executive    |     5 |     2
 Research     |     6 |     2
(5 rows)
```

Department

depid	depname
5	Executive
6	Research
7	Sales
8	Engineering

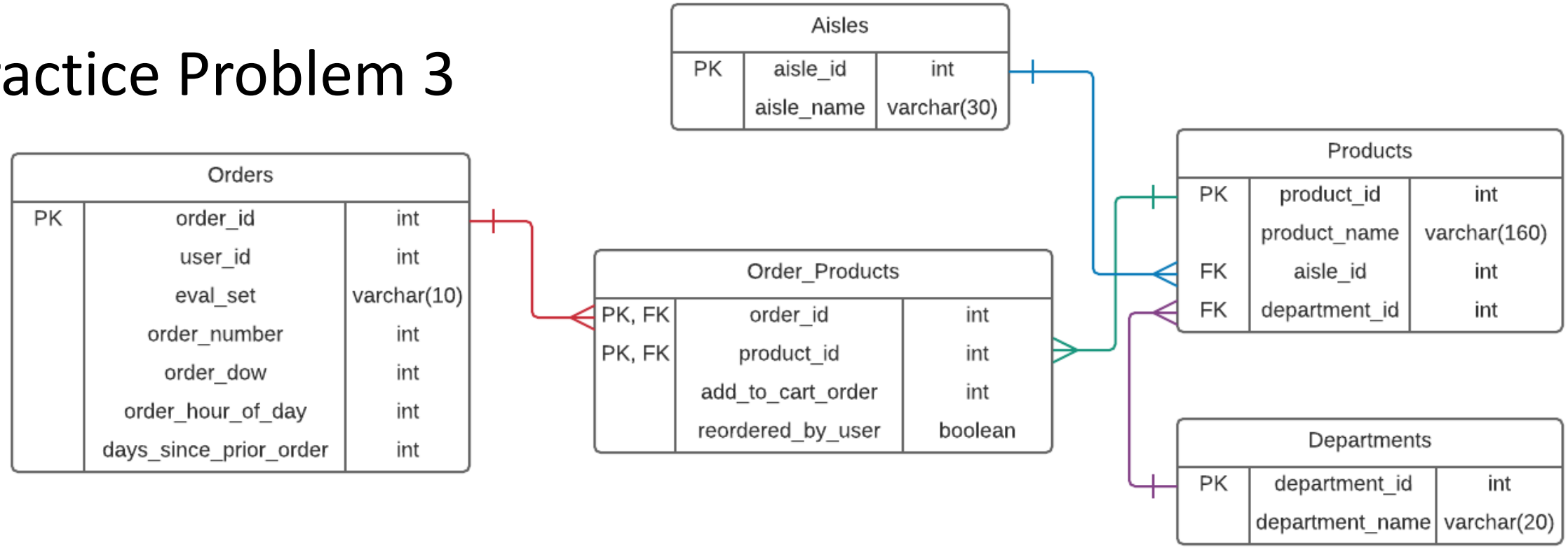
Practice Problem 3



Find the 10 **least** popular products

Display the product name and number of times it was ordered

Practice Problem 3



Find the 10 **least** popular products

Does the query require an outer join?

A) Yes

B) No

C) Unsure

Solutions to Practice Problems:

https://github.com/cs327e-fall2017/snippets/blob/master/instacart_aggregate_queries.sql