# **CSE 344**

#### JUNE 22<sup>ND</sup> INTRODUCTION TO JOINS

(2.1-2.3 & 6.1-6.2)

### REVIEW

#### Data model gives languages for

- describing schema (what data is allowed in the DB)
- writing queries (asking questions & updating data)

#### Relational data model

- database is a collection of tables
- schema describes each table
  - name of table and columns
  - types of all columns
- query language (SQL for now)
  - insert, remove, and print rows of table
  - more to come...

## **ADMINISTRIVIA**

- Should have access to your gitlab repository
- HW1 starter code is there
  - Fill in the .sql files
  - Use test script to check that it works
    - (we will test more thoroughly)
  - Commit, tag, and push to gitlab to turn it in

### **DEMO 1**

- What operations should we expect SQLite (or any DBMS) to support just on what we know right now?
  - create table
  - insert into
  - select
  - delete from

#### • What sorts of inputs do these functions need to have?

- create table: table name, schema
- insert into: table name, tuple
- select: table name, attributes
- delete from: table name, condition

### **DEMO 1**

#### Common Syntax

- CREATE TABLE [tablename] ([att1] [type1], [att2] [type2]...);
- INSERT INTO [tablename] VALUES ([val1],[val2]...);
- SELECT [att1],[att2],... FROM [tablename] WHERE [condition]
- DELETE FROM [tablename]
   WHERE [condition]



- Two other operations we want to support
  - ALTER TABLE: Adds a new attribute to the table
  - UPDATE: Change the attribute for a particular tuple in the table (rather than insert/delete)
- Common Syntax
  - ALTER TABLE [tablename] ADD [attname] [atttype]
  - UPDATE [tablename] SET [attname]=[value]

- Two other operations we want to support
  - ALTER TABLE: Adds a new attribute to the table
  - UPDATE: Change the attribute for a particular tuple in the table (rather than insert/delete)
- Common Syntax
  - ALTER TABLE [tablename] ADD [attname] [atttype]
  - UPDATE [tablename] SET [attname]=[value]
     WHERE [condition]



#### **Tables are NOT ordered**

they are sets or multisets (bags)

#### **Tables are FLAT**

No nested attributes

#### Tables DO NOT prescribe how they are stored on disk

• This is called **physical data independence** 

All three allow DBMSs to be more efficient.

(Last one also simplifies application development.)

• Tables may not be ordered, but data can be returned in an order with the ORDER BY modifier

- Tables may not be ordered, but data can be returned in an order with the ORDER BY modifier
- Whew, today's been a lot of coding... I know what you're thinking...

• We can think of accessing information through queries as some combination of functions

- We can think of accessing information through queries as some combination of functions
  - Consider a table of UW students (with all relevant info):

- We can think of accessing information through queries as some combination of functions
  - Consider a table of UW students (with all relevant info):
    - How would we need to get the birth year of all CSE students from California?

- We can think of accessing information through queries as some combination of functions
  - Consider a table of UW students (with all relevant info):
    - How would we need to get the birth year of all CSE students from California?
    - Think of the file as a set of tuples

- We can think of accessing information through queries as some combination of functions
  - Consider a table of UW students (with all relevant info):
    - How would we need to get the birth year of all CSE students from California?
    - Think of the file as a set of tuples
    - Find the set of CSE students and the set of students from California; Find the intersection of these sets, return just the year from the birthday values of this set

- We can think of accessing information through queries as some combination of functions
  - Consider a table of UW students (with all relevant info):
    - How would we need to get the birth year of all CSE students from California?
    - Think of the file as a set of tuples
    - Find the set of CSE students and the set of students from California; Find the intersection of these sets, return just the year from the birthday values of this set
    - What does this return?

- We can think of accessing information through queries as some combination of functions
  - Consider a table of UW students (with all relevant info):
    - How would we need to get the birth year of all CSE students from California?
    - Think of the file as a set of tuples
    - Find the set of CSE students and the set of students from California; Find the intersection of these sets, return just the year from the birthday values of this set
    - What does this return?
    - Years, but with many duplicates. Even though sets don't allow duplicates, the objects are unique.

- If we only want to return unique elements, we can use the DISTINCT modifier
  - Even if we hide some attributes from the output, the data is all still there.
  - When we select a subset of the attributes, this function is called a *projection* 
    - projections usually produce duplicate values
    - takes work to remove them, so DBMSs usually leave them
    - except on disk, DBMSs work with multisets not sets

- This was all for a single table.
- Data models specify how our data are stored and how the data are related
- Need to utilize these relations, or the database was pointless
- This involves a JOIN

- This was all for a single table.
- Data models specify how our data are stored and how the data are related
- Need to utilize these relations, or the database was pointless
- This involves a JOIN
  - 1NF makes us split up data that belongs together, so query language must make it easy to put them back together whenever necessary
  - we do this with joins

## **JOIN: INTRO**

- The JOIN is the way we use the relationships between tables in a query
  - Example: if we want all of the products and their relevant company information, we need to join those two tables.
  - The result of the join is all of the relevant information from both tables
    - Join occurs based on the join condition.
  - By default, join produces every combination of tuples from the two tables as a row
    - join condition allows you to restrict to the combinations that make sense
    - DBMSs are very good at joining efficiently

#### **JOINS IN SQL**

pname	price	category	manufacturer	cname	country
MultiTouch	199.99	gadget	Canon	GizmoWorks	USA
SingleTouch	49.99	photography	Canon	Canon	Japan
Gizom	50	gadget	GizmoWorks	Hitachi	Japan
SuperGizmo	250.00	gadget	GizmoWorks		-

Retrieve all Japanese products that cost < \$150

#### **JOINS IN SQL**

pname	price	category	manufacturer		cname	country
MultiTouch	199.99	gadget	Canon		GizmoWorks	USA
SingleTouch	49.99	photography	Canon	,	Canon	Japan
Gizom	50	gadget	GizmoWorks		Hitachi	Japan
SuperGizmo	250.00	gadget	GizmoWorks			

Retrieve all Japanese products that cost < \$150

SELECT pname, price FROM Product, Company WHERE ...

#### **JOINS IN SQL**

pname	price	category	manufacturer		cname	country
MultiTouch	199.99	gadget	Canon		GizmoWorks	USA
SingleTouch	49.99	photography	Canon	,	Canon	Japan
Gizom	50	gadget	GizmoWorks		Hitachi	Japan
SuperGizmo	250.00	gadget	GizmoWorks			

Retrieve all Japanese products that cost < \$150

SELECT pname, price
FROM Product, Company
WHERE manufacturer=cname AND
country='Japan' AND price < 150</pre>

#### **JOINS IN SQL**

pname	price	category	manufacturer	cname	country
MultiTouch	199.99	gadget	Canon	GizmoWorks	USA
SingleTouch	49.99	photography	Canon	Canon	Japan
Gizom	50	gadget	GizmoWorks	Hitachi	Japan
SuperGizmo	250.00	gadget	GizmoWorks		-

Re	etrieve	e all Japanese pro Alternative cost < \$150
		syntax
	FROM	pname, price Product JOIN Company manufacturer=cname AND country='Japan' AND price < 150

#### **JOINS IN SQL**

pname	price	category	manufacturerCanonCanonGizmoWorks		cname	country
MultiTouch	199.99	gadget			GizmoWorks	USA
SingleTouch	49.99	photography			Canon	Japan
Gizom	50	gadget			Hitachi	Japan
SuperGizmo	250.00	gadget	GizmoWorks	L		

Retrieve all Japanese products that cost Alternative

syntax

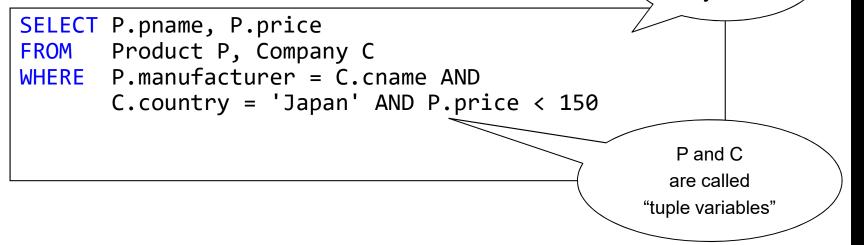
SELECT pname, price FROM Product JOIN Company ON manufacturer=cname WHERE country='Japan' AND price < 150

#### **JOINS IN SQL**

pname	price	category	manufacturer	cname	country
MultiTouch	199.99	gadget	Canon	GizmoWorks	USA
SingleTouch	49.99	photography	Canon	Canon	Japan
Gizom	50	gadget	GizmoWorks	Hitachi	Japan
SuperGizmo	250.00	gadget	GizmoWorks		•

Retrieve all Japanese products that cost Alternative

syntax



#### **JOINS IN SQL**

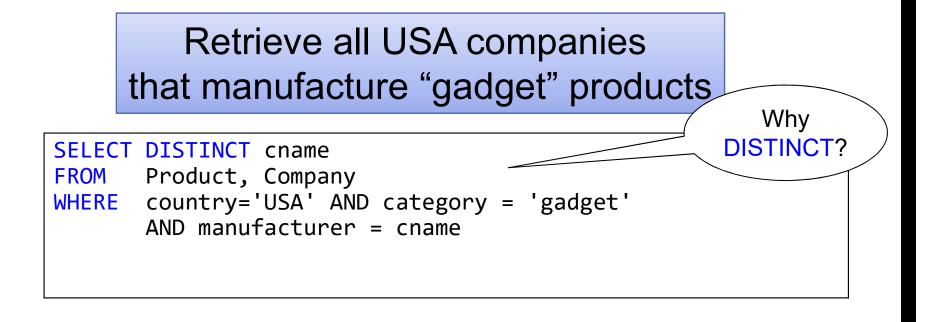
pname	price	category	manufacturer	cname
MultiTouch	199.99	gadget	Canon	GizmoW
SingleTouch	49.99	photography	Canon	Canon
Gizom	50	gadget	GizmoWorks	Hitachi
SuperGizmo	250.00	gadget	GizmoWorks	

cname	country
GizmoWorks	USA
Canon	Japan
Hitachi	Japan

Retrieve all USA companies that manufacture "gadget" products

#### **JOINS IN SQL**

pname	price	category	manufacturer		cname	country
MultiTouch	199.99	gadget	Canon		GizmoWorks	USA
SingleTouch	49.99	photography	Canon		Canon	Japan
Gizom	50	gadget	GizmoWorks		Hitachi	Japan
SuperGizmo	250.00	gadget	GizmoWorks			· · ·



## **JOINS IN SQL**

The standard join in SQL is called an inner join

Each row in the result must come from both tables in the join

Sometimes we want to include rows from only one of the two table: outer join

Employee(id, name)
Sales(employeeID, productID)

## **INNER JOIN**

Employee			Sales							
id	name		employeeID	productID						
1	Joe		1	344						
2	Jack		1	355						
3	Jill		2	544						

Retrieve employees and their sales

Employee(id, name)
Sales(employeeID, productID)

## **INNER JOIN**

Employee			Sales						
id	name		<u>employeeID</u>	productID					
1	Joe		1	344					
2	Jack		1	355					
3	Jill		2	544					

Retrieve employees and their sales

SELECT \*
FROM Employee E, Sales S
WHERE E.id = S.employeeID

Employee(id, name)
Sales(employeeID, productID)

## **INNER JOIN**

Employee		Sales		
id	name	<u>employeeID</u>	productID	
1	Joe	1	344	
2	Jack	1	355	
3	Jill	2	544	

#### Retrieve employees and their sales

SELECT	*
FROM	Employee E, Sales S
WHERE	E.id = S.employeeID
	. 2

id	name	employeeID	productID
1	Joe	1	344
1	Joe	1	355
2	Jack	2	544

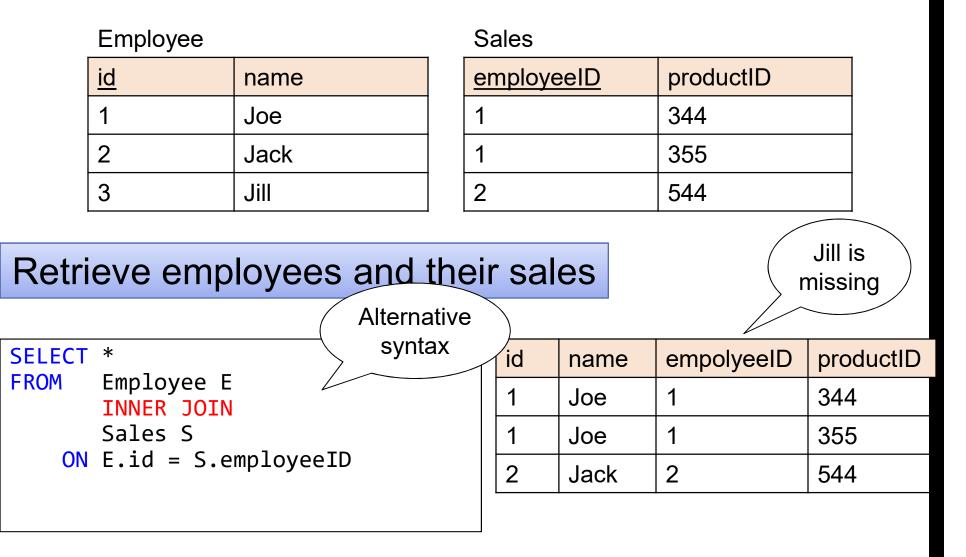
Employee(id, name)
Sales(employeeID, productID)

# **INNER JOIN**

	Employee		5	Sales			
	id	name	e	employe	<u>eeID</u>	productID	
	1	Joe	-	1		344	
	2	Jack	-	1		355	
	3	Jill		2		544	
Retrieve employees and		the	ir sal	es	r	Jill is nissing	
SELECT				id	name	employeeID	productID
	Employee E E.id = S.e	-		1	Joe	1	344
			1	Joe	1	355	

Employee(id, name)
Sales(employeeID, productID)

# **INNER JOIN**



Employee(id, name)
Sales(employeeID, productID)

# **OUTER JOIN**

	Employee			Sa	ales			
	id	name		er	nploye	<u>eeID</u>	productID	
	1	Joe		1			344	
	2	Jack		1			355	
	3	Jill		2			544	
Retr	ieve emp	loyees and	the	eir	r sal	es	Jill is presen	t
SELECT					id	name	empolyeeID	productID
FROM	Employee E				1	Joe	1	344
	Sales S				1	Joe	1	355
	E.id = S.e	mpioyeeiD			2	Jack	2	544
					3	Jill	NULL	NULL

# (INNER) JOINS

```
Product(pname, price, category, manufacturer)
Company(cname, country)
-- manufacturer is foreign key to Company
```

```
SELECT DISTINCT cname
FROM Product, Company
WHERE country='USA' AND category = 'gadget'
AND manufacturer = cname
```

#### Product

pname	category	manufacturer
Gizmo	gadget	GizmoWorks
Camera	Photo	Hitachi
OneClick	Photo	Hitachi

cname	country
GizmoWorks	USA
Canon	Japan
Hitachi	Japan

#### Product

pname	category	manufacturer	
Gizmo	gadget	GizmoWorks	
Camera	Photo	Hitachi	
OneClick	Photo	Hitachi	

cname	country
GizmoWorks	USA
Canon	Japan
Hitachi	Japan

#### Product

pname	category	manufacturer
Gizmo	gadget	GizmoWorks
Camera	Photo	Hitachi
OneClick	Photo	Hitachi

cname	country
GizmoWorks	USA
Canon	Japan
Hitachi	Japan

Product

pname	category	manufacturer	 cname	country
Gizmo	gadget	GizmoWorks	GizmoWorks	USA
Camera	Photo	Hitachi	 Canon	Japan
OneClick	Photo	Hitachi	Hitachi	Japan

pname	category	manufacturer	cname	country
Gizmo	gadget	GizmoWorks	GizmoWorks	USA

#### Product

pname	category	manufacturer	
Gizmo	gadget	GizmoWorks	
Camera	Photo	Hitachi	
OneClick	Photo	Hitachi	

cname	country
GizmoWorks	USA
Canon	Japan
Hitachi	Japan

#### Product

pname	category	manufacturer	
Gizmo	gadget	GizmoWorks	
Camera	Photo	Hitachi	
OneClick	Photo	Hitachi	

cname	country
GizmoWorks	USA
Canon	Japan
Hitachi	Japan

SELECT	DISTINCT cname
FROM	Product JOIN Company ON
	country = 'USA' AND
	category = 'gadget' AND
	manufacturer = cname

(INNER) JOINS SELECT x1.a1, x2.a2, ... xm.am FROM R1 as x1, R2 as x2, ... Rm as xm WHERE Cond

Product(pname, price, category, manufacturer)
Company(cname, country)

-- manufacturer is foreign key to Company

```
Product(<u>pname</u>, price, category, manufacturer)
Company(<u>cname</u>, country)
```

-- manufacturer is foreign key to Company

```
SELECT DISTINCT z.cname
FROM Product x, Company z
WHERE z.country = 'USA'
AND x.manufacturer = z.cname
AND x.category = 'gadget'
AND x.category = 'photography;
```



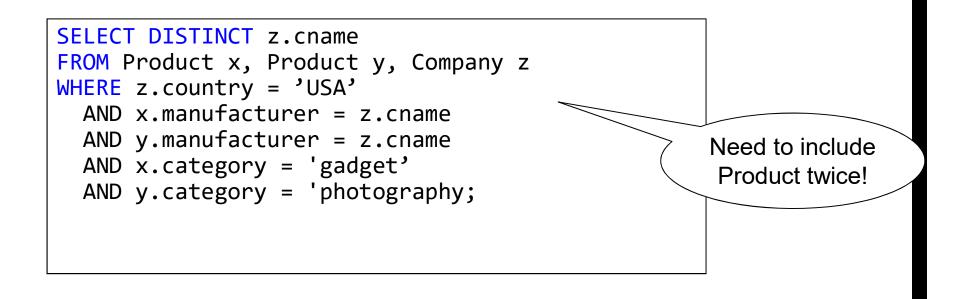
```
Product(pname, price, category, manufacturer)
Company(cname, country)
```

-- manufacturer is foreign key to Company



```
Product(pname, price, category, manufacturer)
Company(cname, country)
```

-- manufacturer is foreign key to Company



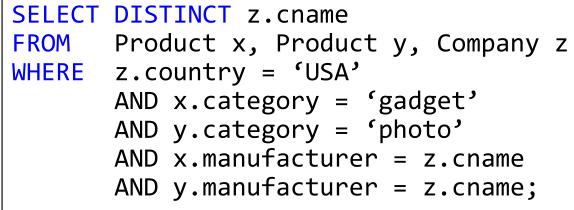
# **SELF-JOINS AND TUPLE VARIABLES**

Find USA companies that manufacture both products in the 'gadgets' and 'photo' category

Joining Product with Company is insufficient: need to join Product, with Product, and with Company

When a relation occurs twice in the FROM clause we call it a self-join; in that case we must use tuple variables (why?)

## SELE IUNS

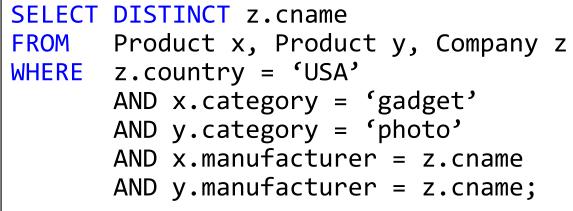


#### Product

pname	category	manufacturer
Gizmo	gadget	GizmoWorks
SingleTouch	photo	Hitachi
MultiTouch	Photo	GizmoWorks

cname	country
GizmoWorks	USA
Hitachi	Japan

## SELE INING



Product

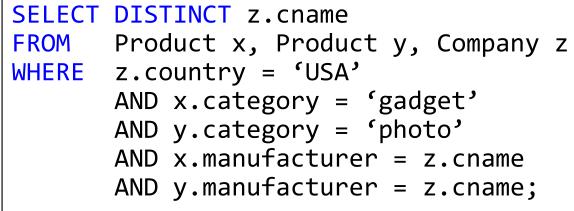
Χ

Com	pany
-----	------

pname	category	manufacturer
Gizmo	gadget	GizmoWorks
SingleTouch	photo	Hitachi
MultiTouch	Photo	GizmoWorks

cname	country
GizmoWorks	USA
Hitachi	Japan

## SELE IUNS

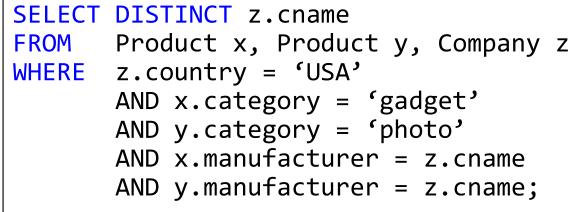


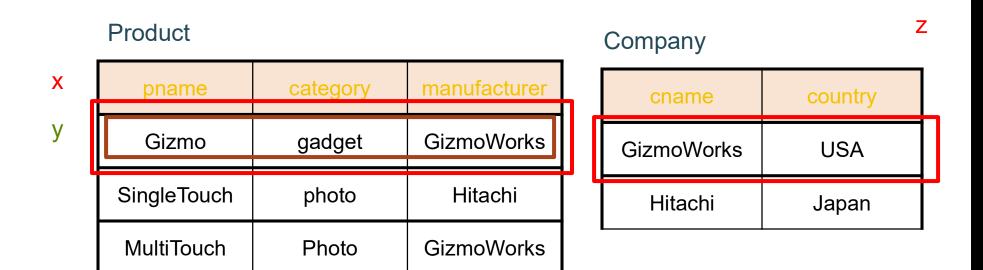
Product

X	pname	category	manufacturer
у	Gizmo	gadget	GizmoWorks
	SingleTouch	photo	Hitachi
	MultiTouch	Photo	GizmoWorks

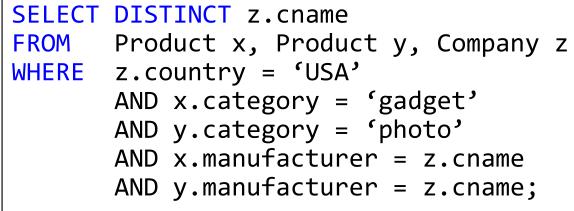
cname	country
GizmoWorks	USA
Hitachi	Japan

## SELE IOINS





## SELE IUNS



Product

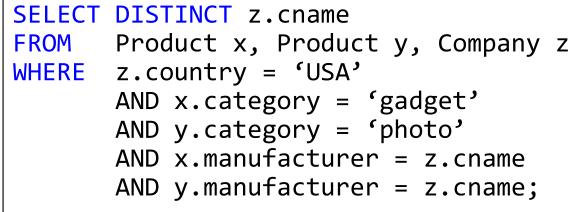
x	pname	category	manufacturer
у	Gizmo	gadget	GizmoWorks
	SingleTouch	photo	Hitachi

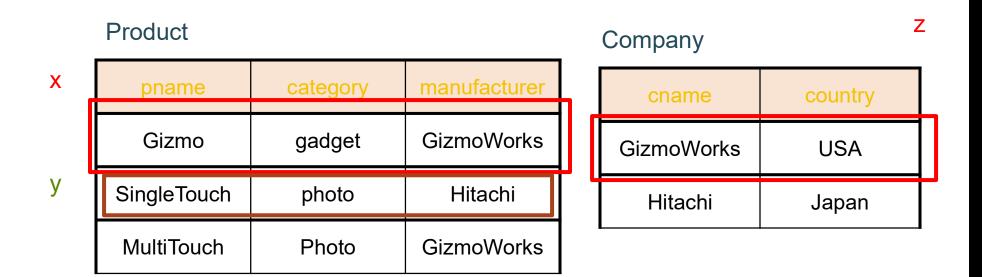
## Company

cname	country	
GizmoWorks	USA	
Hitachi	Japan	

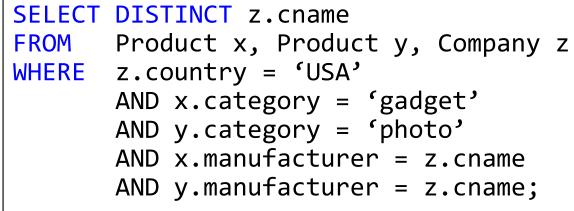
#### Ζ

## SELE IOINS





## SELE IUNS



XpnamecategorymanufacturerGizmogadgetGizmoWorksySingleTouchphotoHitachiMultiTouchPhotoGizmoWorks

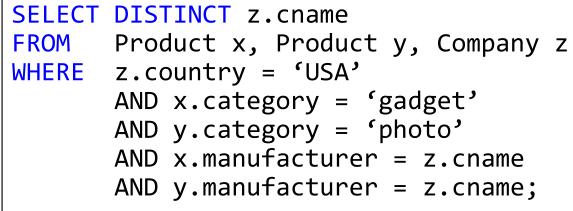
# CompanycnamecountryGizmoWorksUSA

Japan

Hitachi

Ζ

## SELE- INING



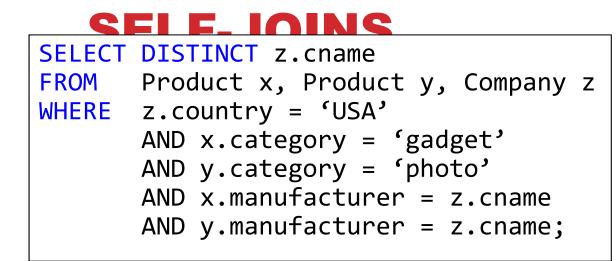
Product

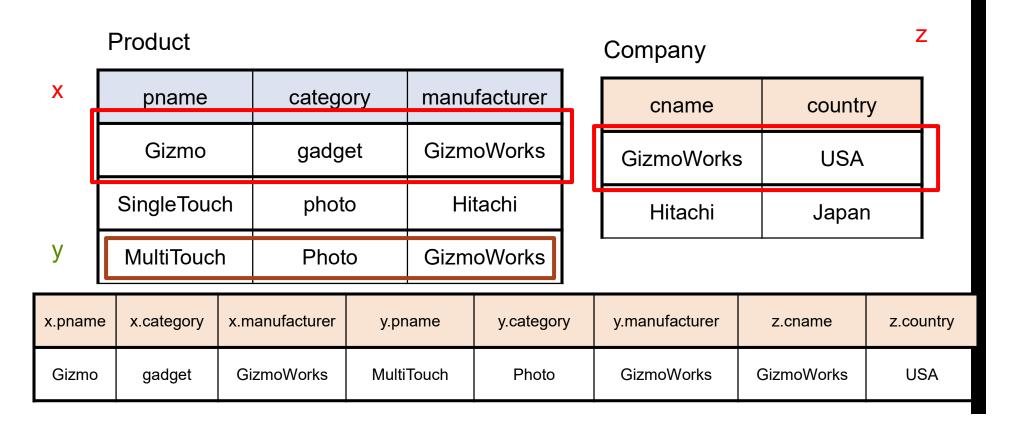
## Company

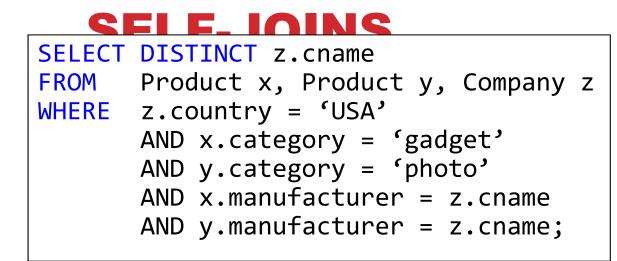


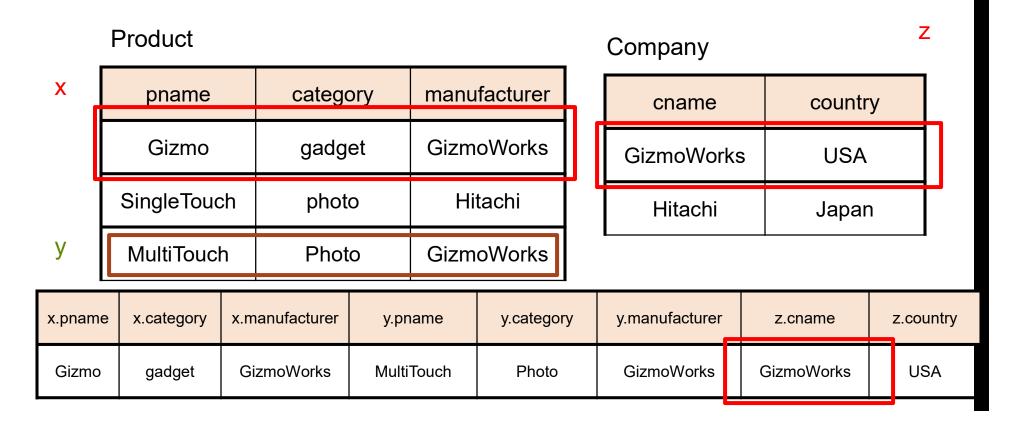
cname	country	
GizmoWorks	USA	
Hitachi	Japan	

Ζ









# **OUTER JOINS**

Product(<u>name</u>, category)
Purchase(prodName, store)

-- prodName is foreign key

SELECT Product.name, Purchase.store
FROM Product, Purchase
WHERE Product.name = Purchase.prodName

We want to include products that are never sold, but some are not listed! Why?

# **OUTER JOINS**

Product(<u>name</u>, category)
Purchase(prodName, store)

-- prodName is foreign key

SELECT Product.name, Purchase.store
FROM Product LEFT OUTER JOIN Purchase ON
Product.name = Purchase.prodName

## SELECT Product.name, Purchase.store FROM Product JOIN Purchase ON Product.name = Purchase.prodName

Product

#### Purchase

Name	Category
Gizmo	gadget
Camera	Photo
OneClick	Photo

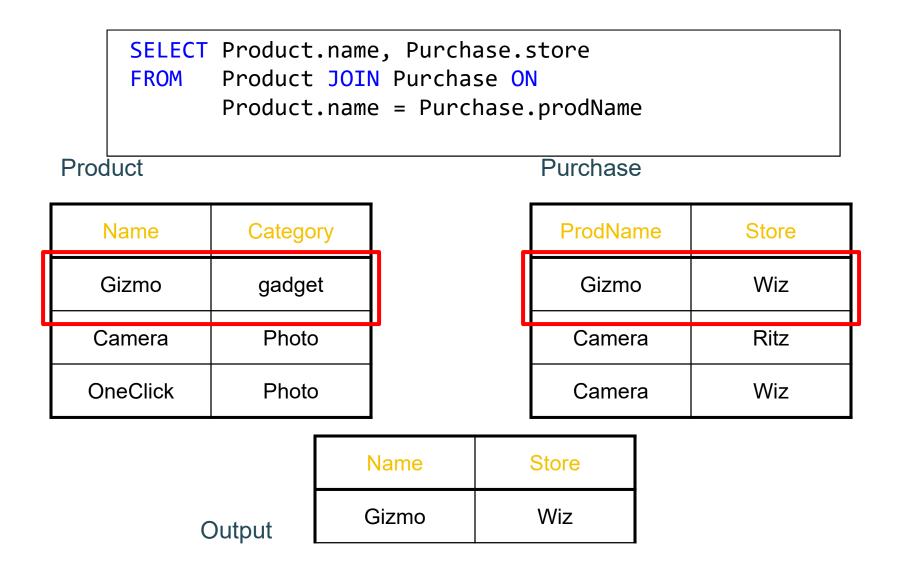
ProdName	Store
Gizmo	Wiz
Camera	Ritz
Camera	Wiz

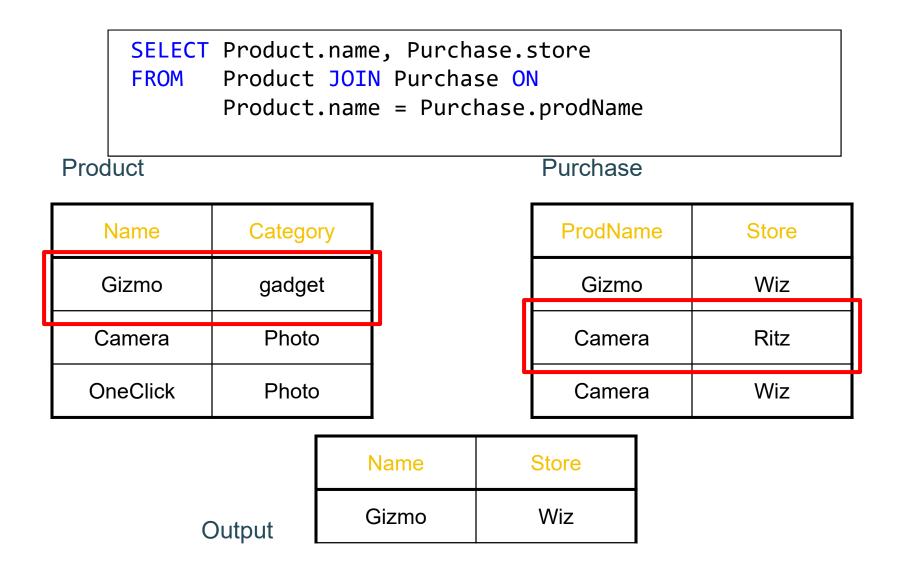
## SELECT Product.name, Purchase.store FROM Product JOIN Purchase ON Product.name = Purchase.prodName

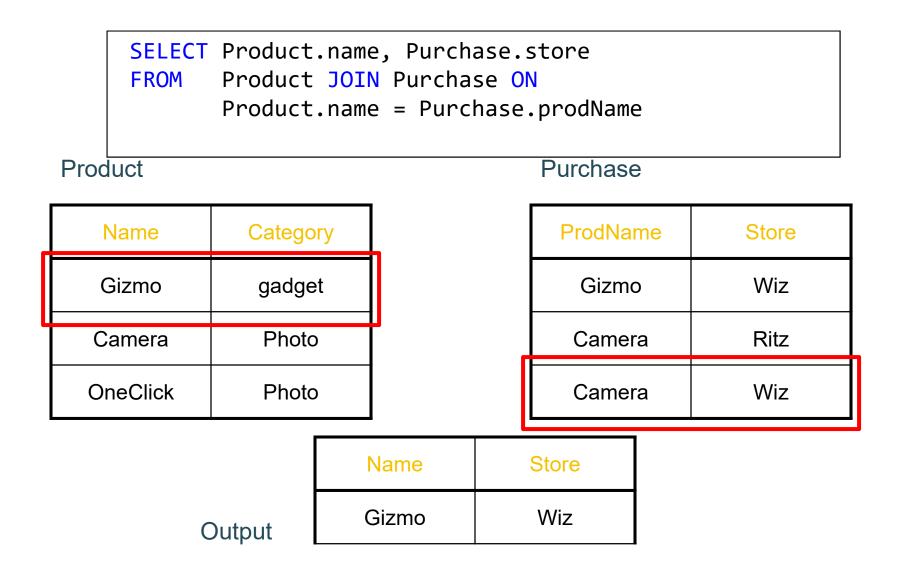
Product

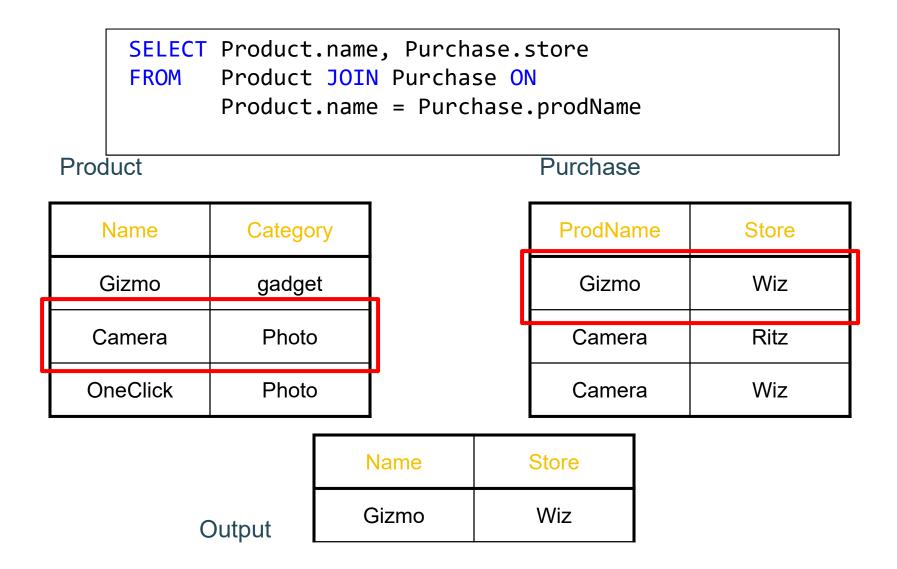
Name	Category
Gizmo	gadget
Camera	Photo
OneClick	Photo

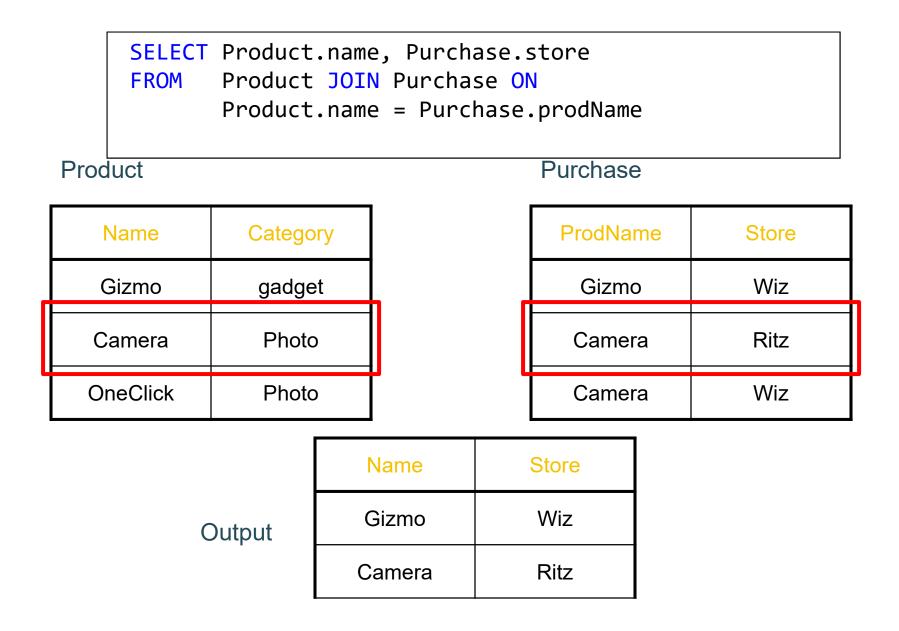
ProdName	Store
Gizmo	Wiz
Camera	Ritz
Camera	Wiz

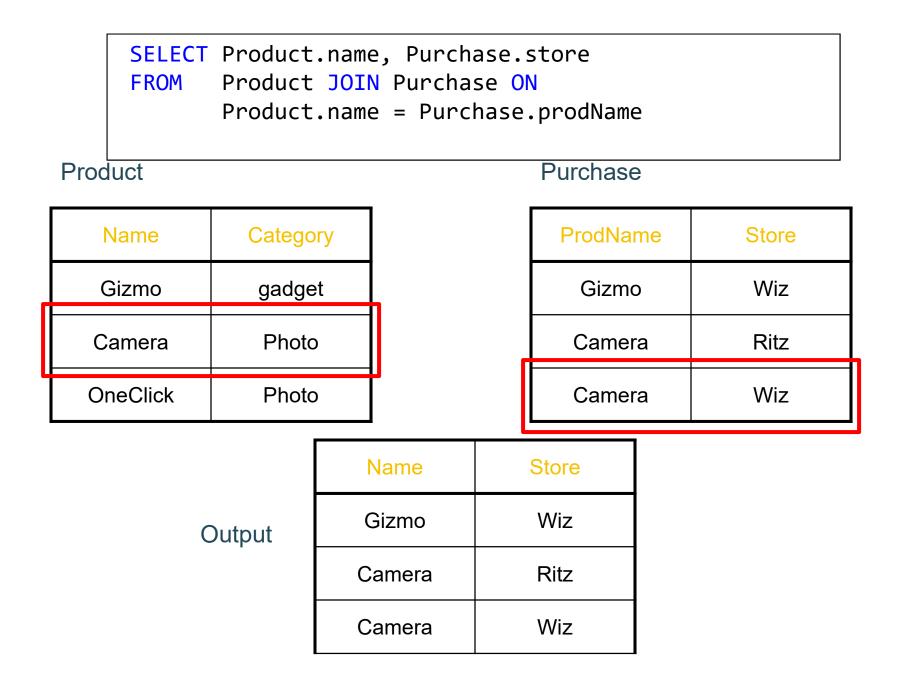


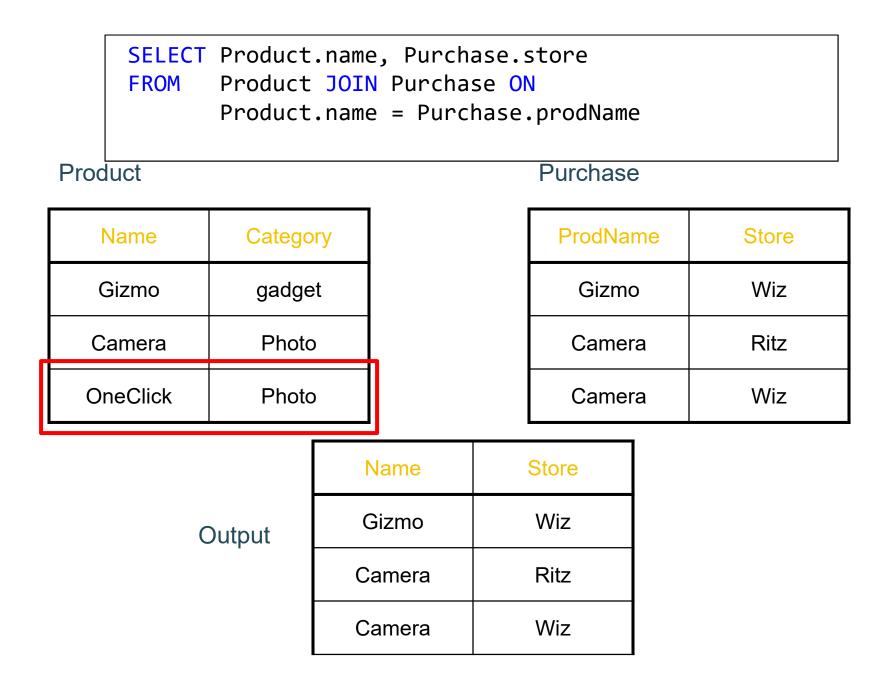


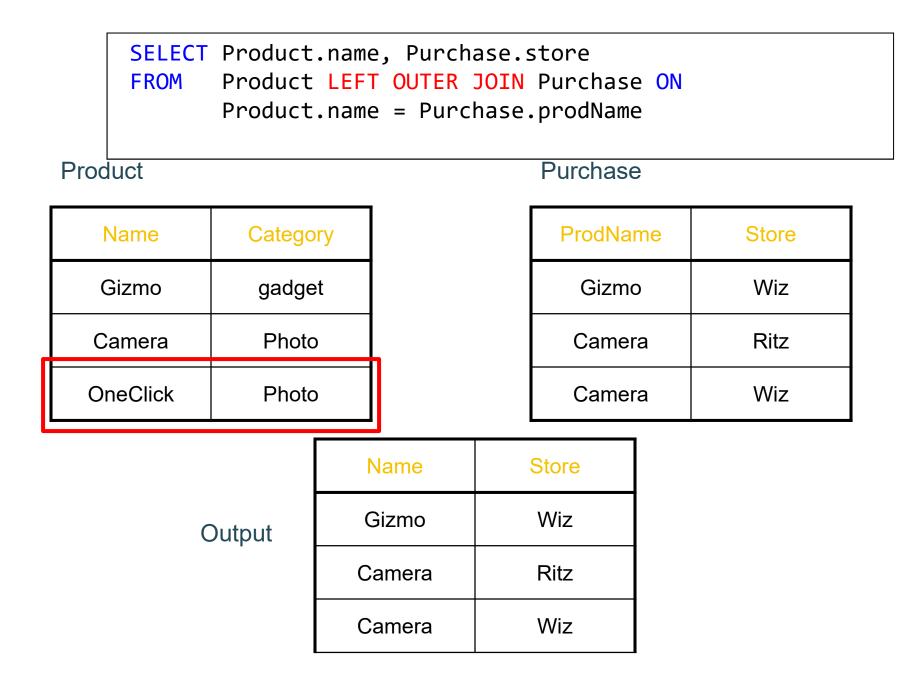


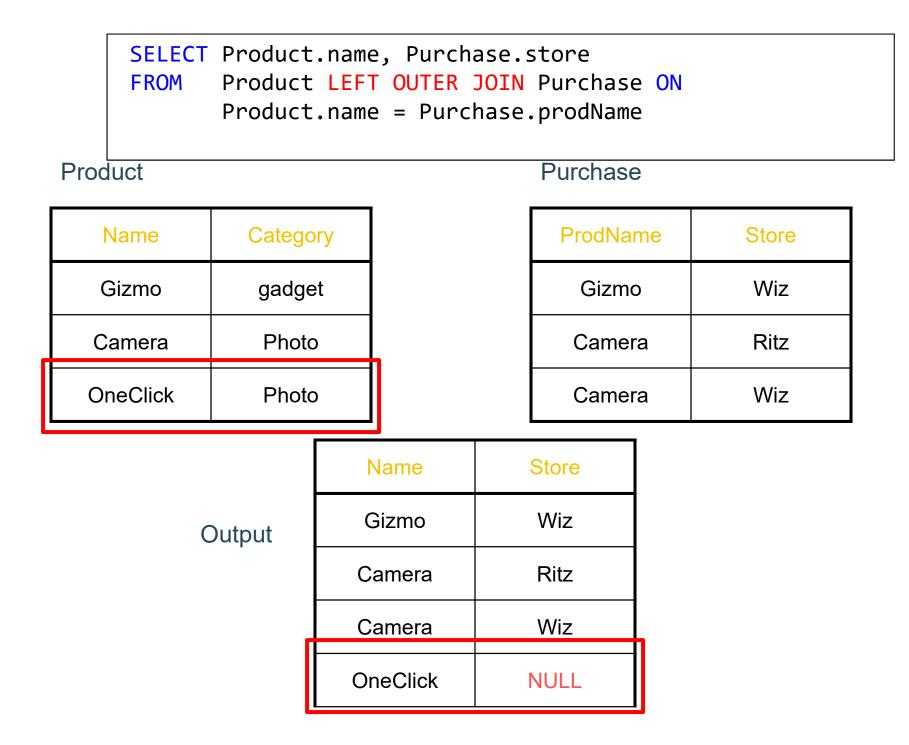


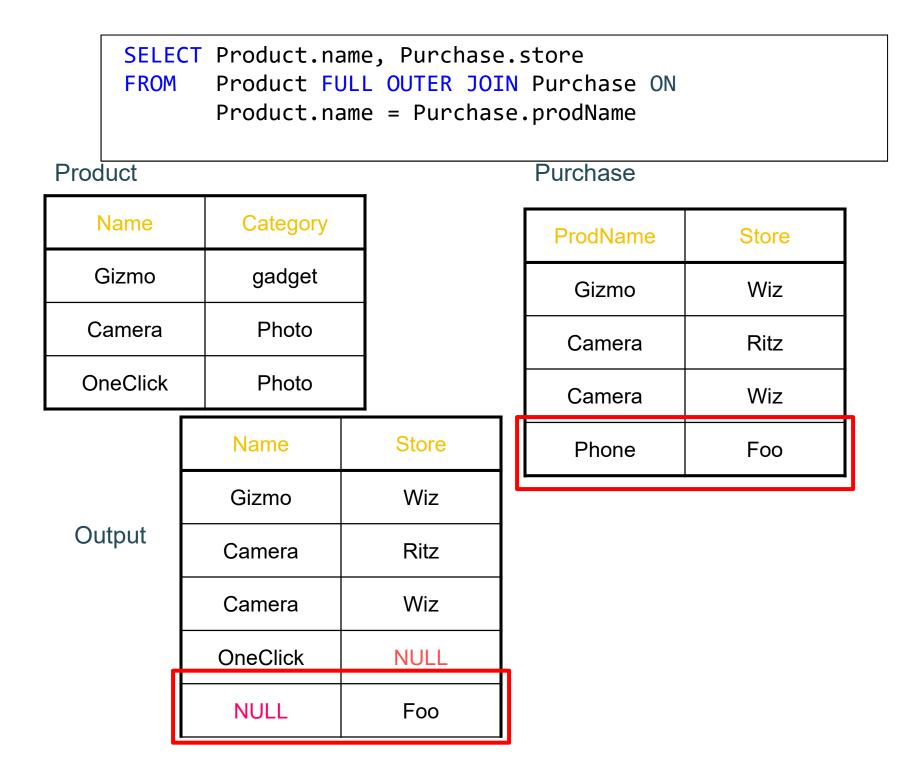












# **OUTER JOINS**

tableA (LEFT/RIGHT/FULL) OUTER JOIN tableB ON p

# Left outer join:

Include tuples from tableA even if no match

# **Right outer join:**

- Include tuples from tableB even if no match
   Full outer join:
  - Include tuples from both even if no match

# In all cases:

Patch tuples without matches using NULL