



### **Agenda**

- > Setup Grakn
- > What is a graph?
- > Why use graphs?
- > Are hypergraphs useful?
- > How do I model a graph?
- > Introducing Grakn
  - Ontology & Knowledge Model
  - Hierarchies & Relations
  - Modeling Tips
- > Introducing Graql
  - Defining Schemas
  - Writing Data
- Reading Data
- Deleting/Modifying Data
- > Exercises

# Setup Grakn

### Setup Grakn & Grakn Workbase

- Download grakn-core-all-\* from <a href="https://github.com/graknlabs/grakn/releases/tag/1.6.2">https://github.com/graknlabs/grakn/releases/tag/1.6.2</a>
- Extract grakn-core-all, open terminal/CMD, and CD into extracted directory
- Run ./grakn server start (or .\grakn.bat server start on Windows)
- Download grakn-workbase-\* from <a href="https://github.com/graknlabs/workbase/releases/tag/1.2.7">https://github.com/graknlabs/workbase/releases/tag/1.2.7</a>
- Extract grakn-workbase, open terminal, and CD into extracted directory
  - On Windows, double-click the .exe to install
- Run ./grakn-workbase
  - On Windows, run installed Grakn Workbase app

### Download/Clone Training Data

https://github.com/BFergerson/sql-to-graql

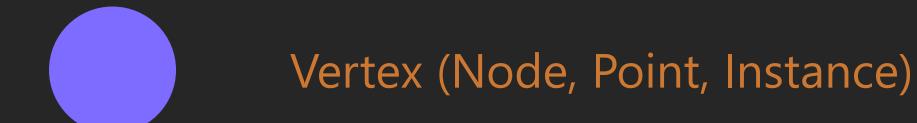
### Executing Graql Files

- On Linux/mac:
  - ./grakn console -f /path/to/file.gql
- On Windows:
  - .\grakn.bat console -f C:\path\to\file.gql

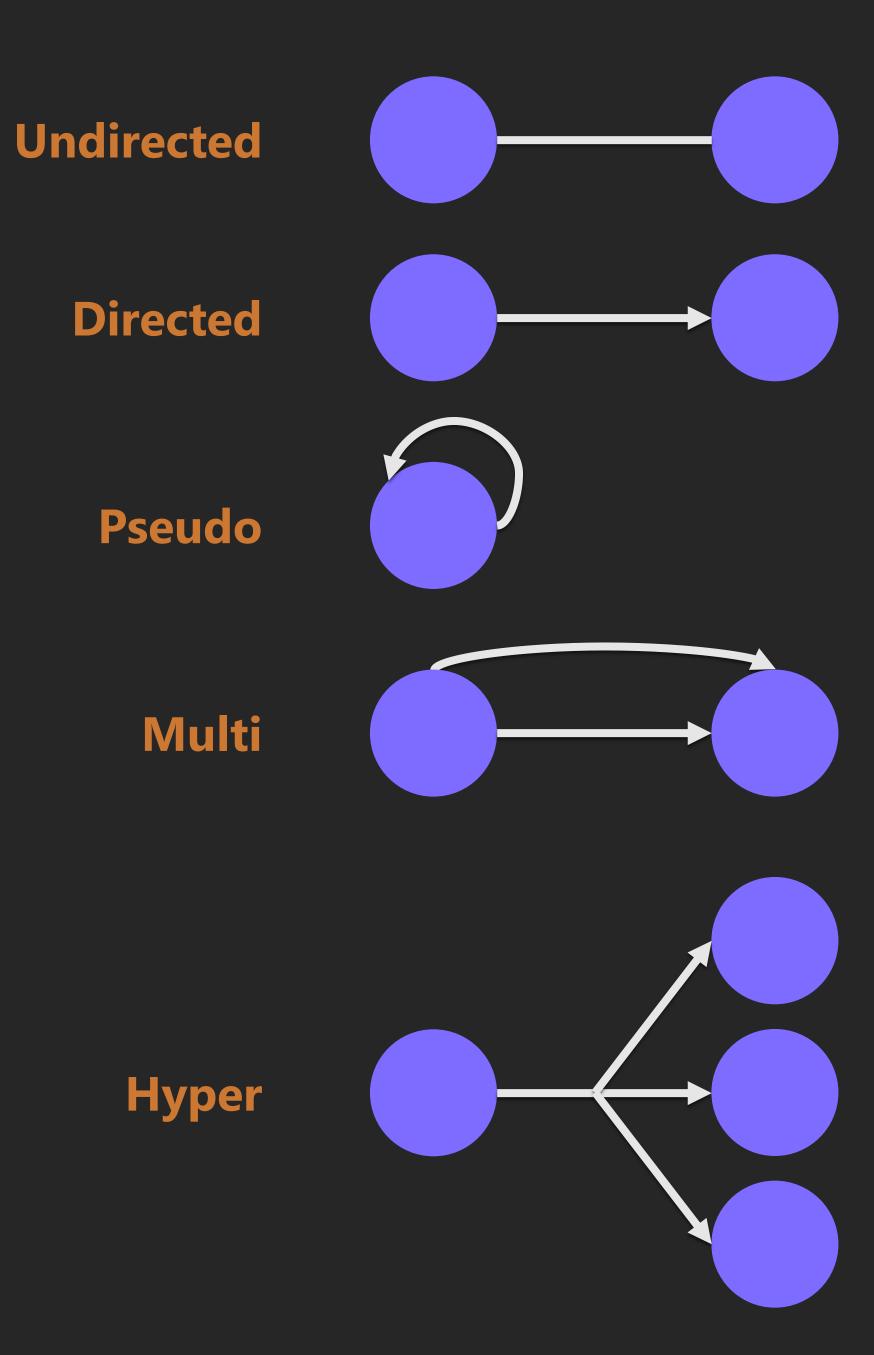
#### Notice:

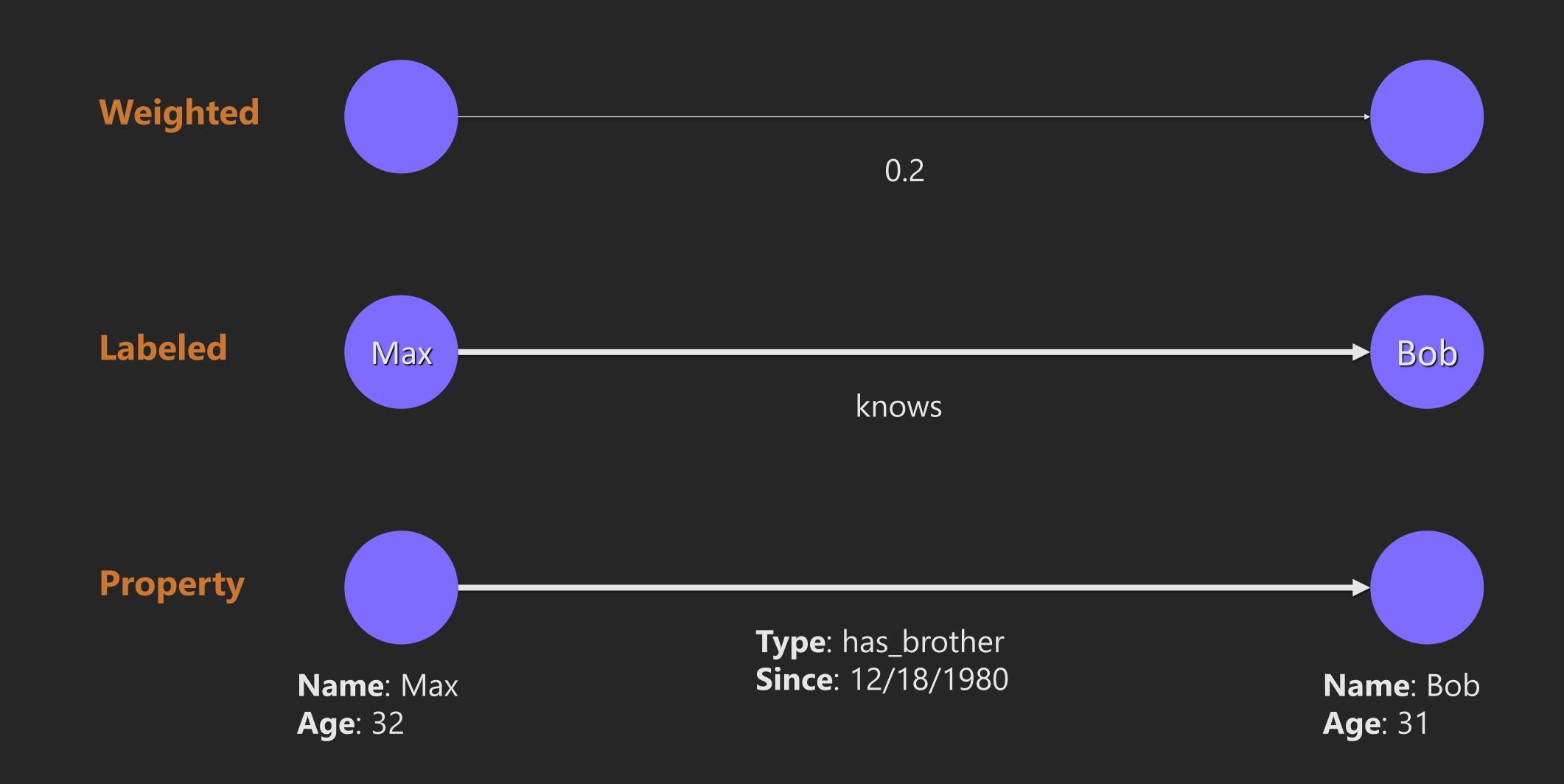
- Files under "graql" are standalone and can be executed separately or together in any order
- Files under "answers/exercise" are standalone and can be executed separately or together in any order
- Files under "answers/fill-in-the-blank" must be executed in the order they appear in this presentation

An abstract representation of a set of objects where some pairs are connected by links.



Edge (Link, Line, Arc, Relation)

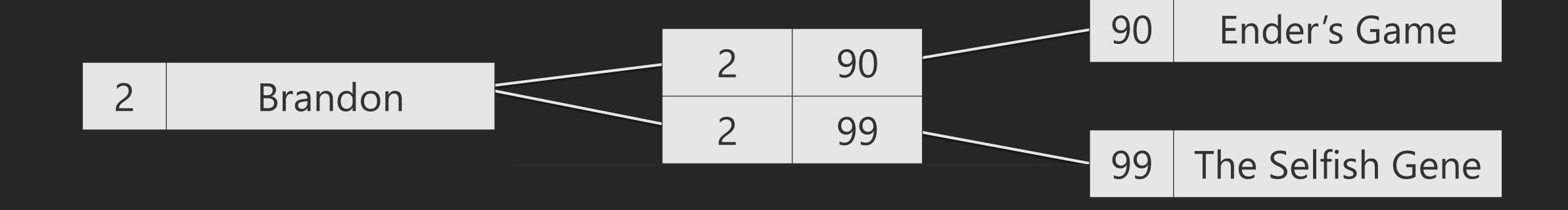


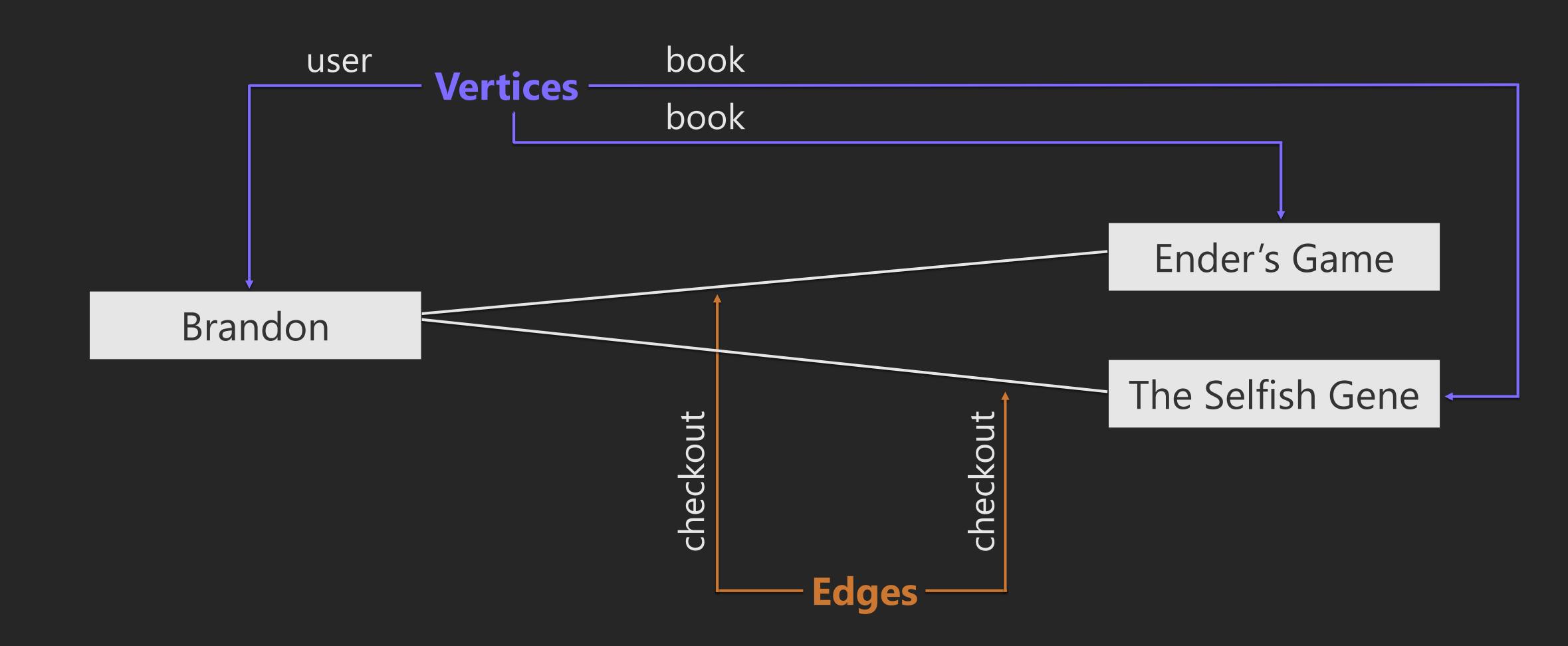


Users		
• • •	• • •	
2	Brandon	
• • •	• • •	

	Checkouts				
	• • •	• • •			
_	2	90			
_	2	99			
	• • •	• • •			

Books				
90	Ender's Game			
• • •	• • •			
99	The Selfish Gene			





#### 1:1 Relationship

				Id	Location
Id	Name	Address		3	Tampa
1	Bob	3		4	Liverpool
2	Max	5	<b>→</b>	5	Vancouver
3	Jeffrey	10		10	
				10	Dubai

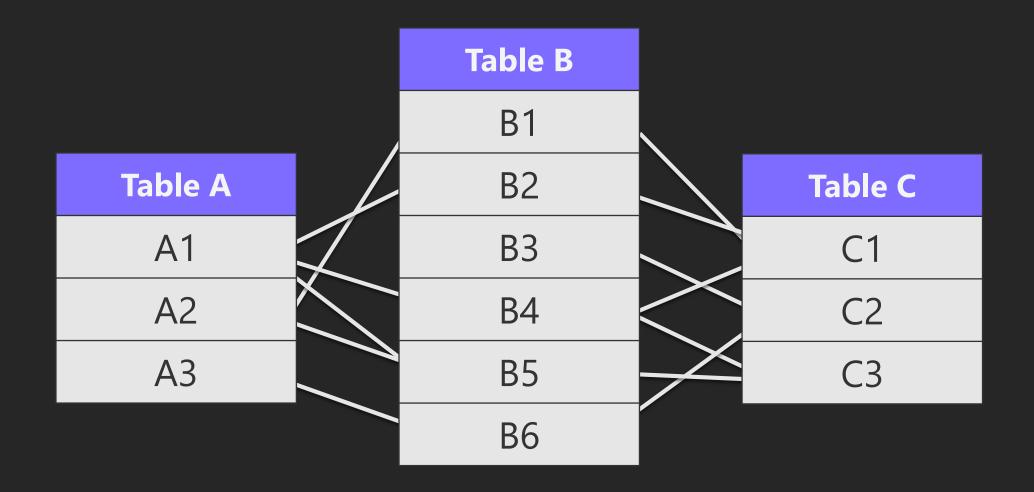
#### 1:n Relationship

			Id	Customer	Location
Id	Name		3	1	Tampa
1	Bob	-	5	2	Vancouver
2	Max		8	2	New York
3	Jeffrey	<b>←</b>		_	
			10	3	Dubai

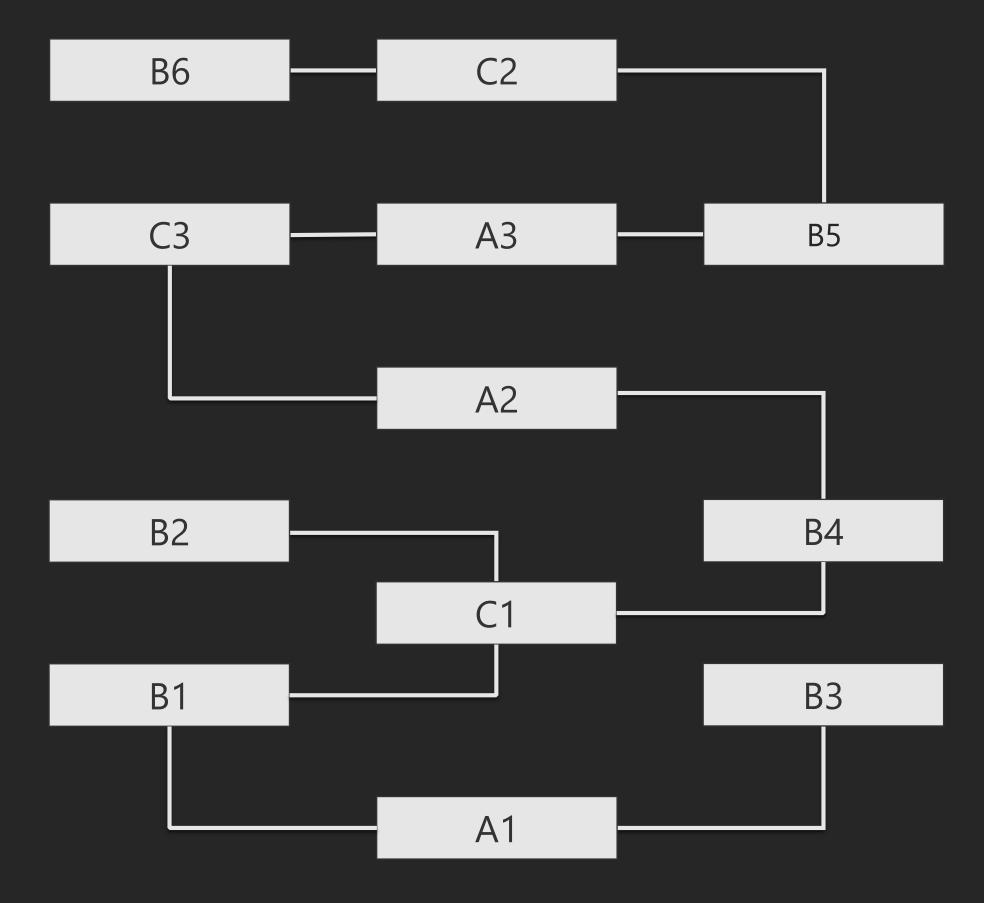
#### m:n Relationship

			Cid	Aid		Id	Location
Id	Name	4	1	3		3	Tampa
1	Bob	4	2	5	<b></b>	5	Vancouver
2	Max	<b>←</b>	2	8	<b></b>	8	New York
3	Jeffrey		3	10		10	Dubai

#### **Optimized for aggregate data**



#### **Optimized for connected data**



#### Use graphs when:

- Shape of the domain is naturally a graph
- Problems with join performance
- Joining more than 7 tables together
- Majority of your tables are junction tables
- Written stored procedures with multiple recursive self and inner joins
- Continuously evolving data set (often means creating wide and sparse tables)
- Constantly changing schema
- Require combing multiple schemas

#### **Graphs are good at:**

- Path finding (how do people know each other)
- Link analysis (fraud detection)
- Highly connected data (social networks)
- Recommendations (e-commerce)
- A\* (least-cost path analysis)

#### **Graphs are designed to:**

- Store interconnected data
- Make it easy to make sense of data connections
- Make it easy to evolve the database
- Enable high-performance on operations for:
  - Discovery of connected data patterns
  - Relatedness queries > depth 1

Graphs RDBMS Document Column Family Key-Value

Size

Friends-of-friends
1,000,000 people
~50 friends each

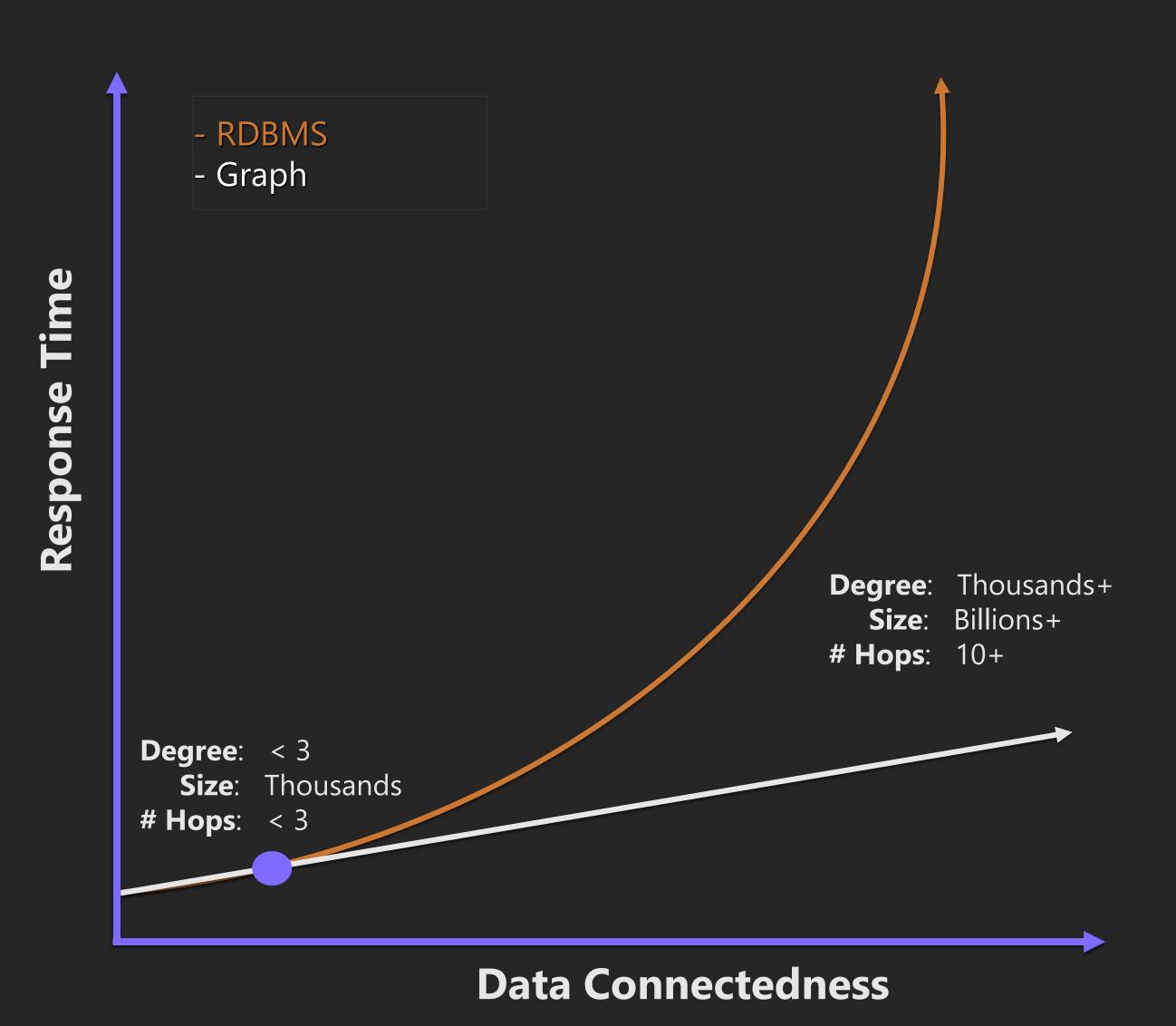
Depth	RDBMS execution time (s)	Neo4j execution time (s)	Records returned
2	0.016	0.01	~2500
3	30.267	0.168	~110k
4	1543.505	1.359	~600k
5	Unfinished	2.132	~800k

Friends-of-friends ~50 friends each Depths of 4

Database	# persons	Query time
MySQL	1,000	2 sec
Neo4j	1,000	2 ms
Neo4j	1,000,000	2 ms

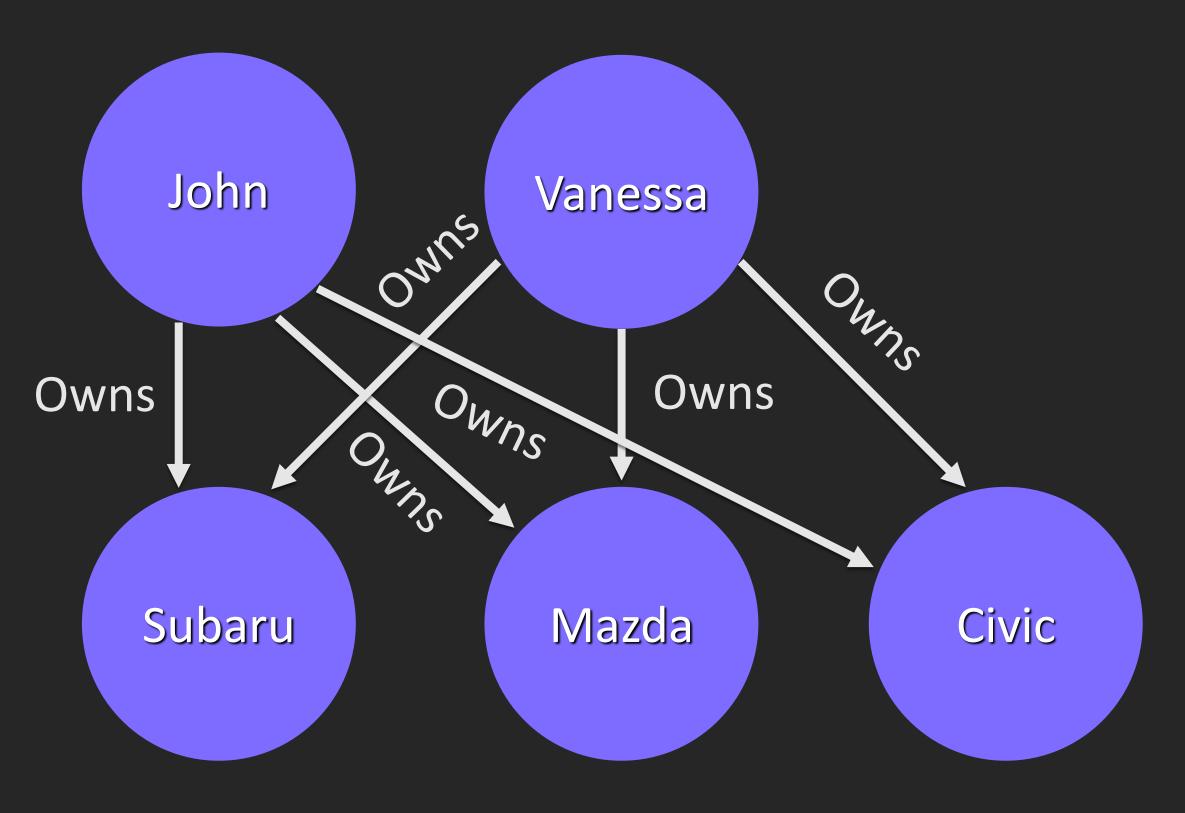
#### **Query Response Time**

- = **f**(graph density, graph size, query degree)
- Graph density (avg # rel's / node)
- Graph size (total # of nodes in the graph)
- Query degree (# of hops in one's query)

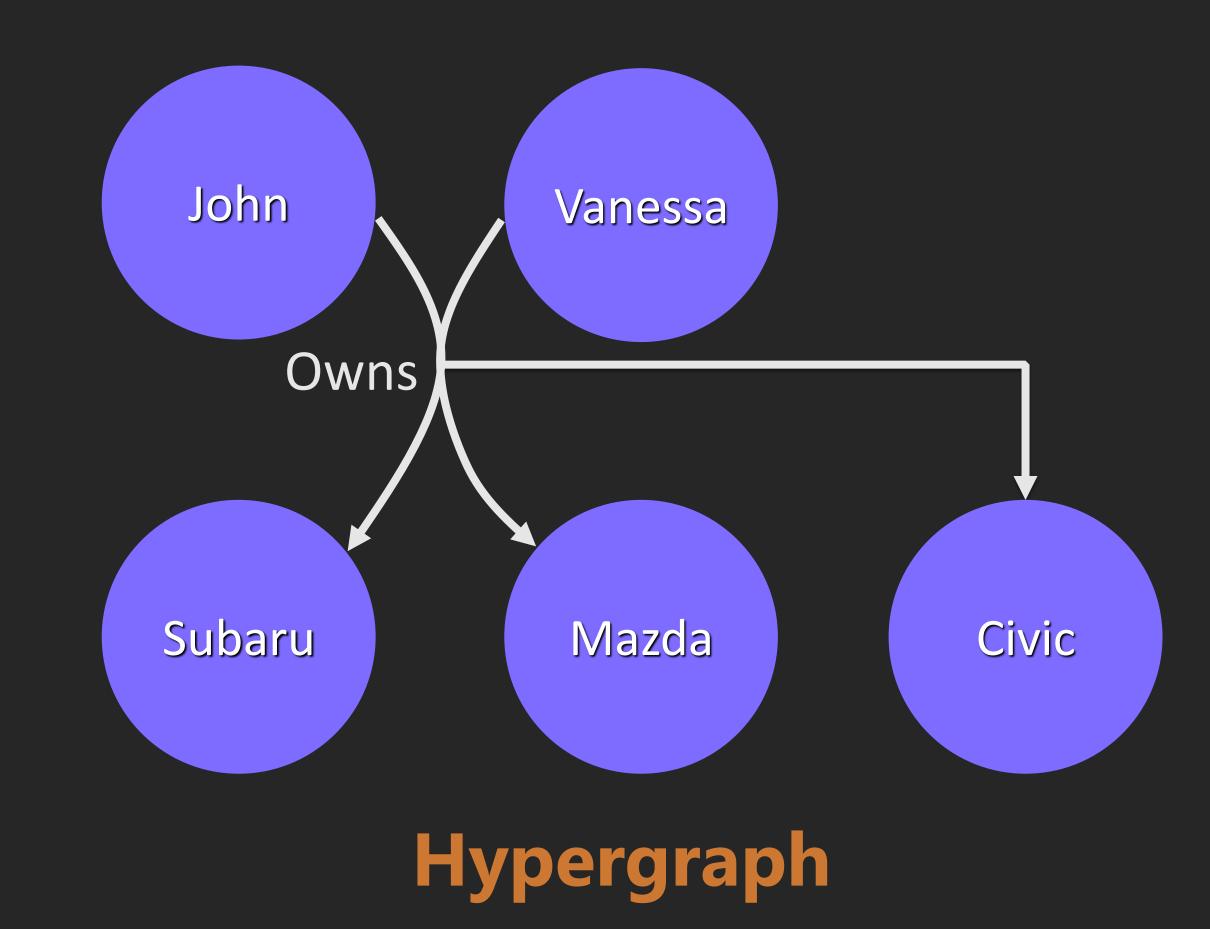


# Are hypergraphs useful?

### Are hypergraphs useful?

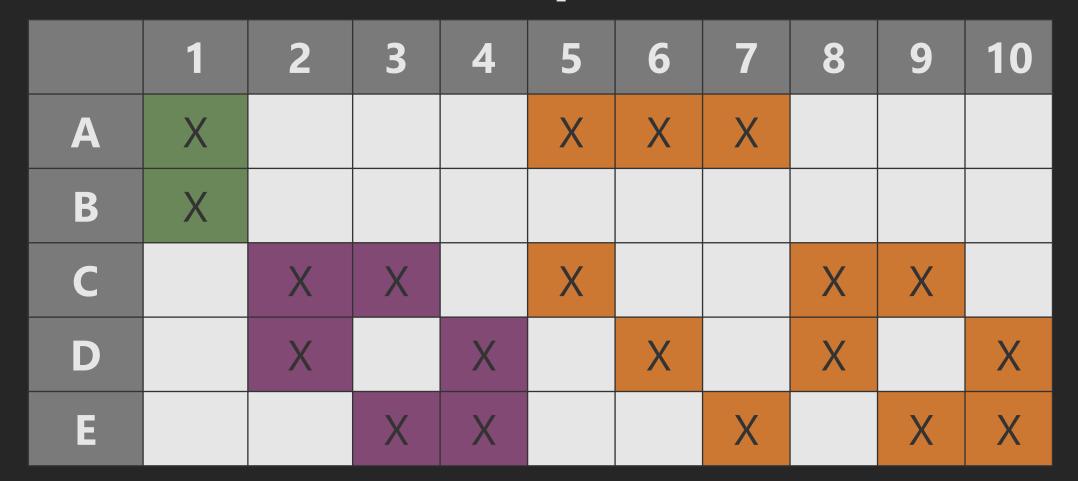


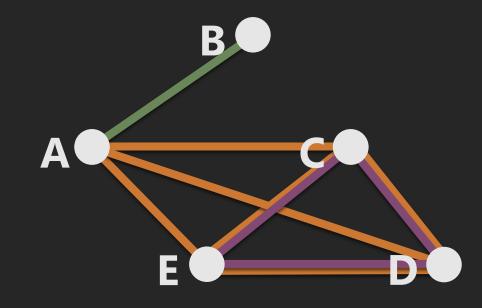
Directed Graph



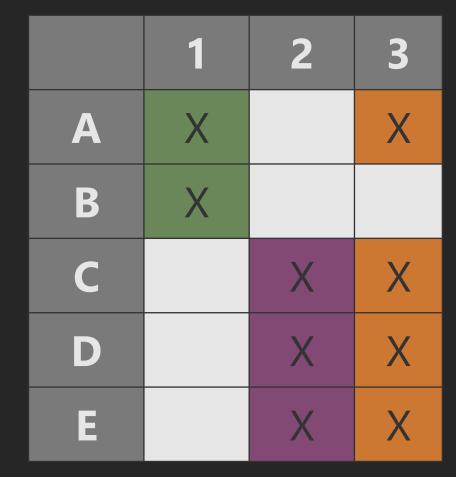
### Are hypergraphs useful?

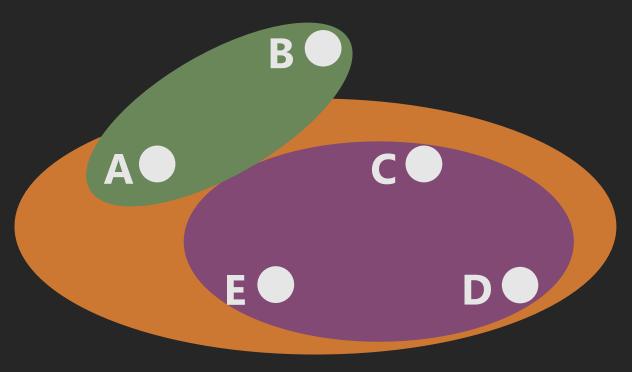
### Graph





#### Hypergraph





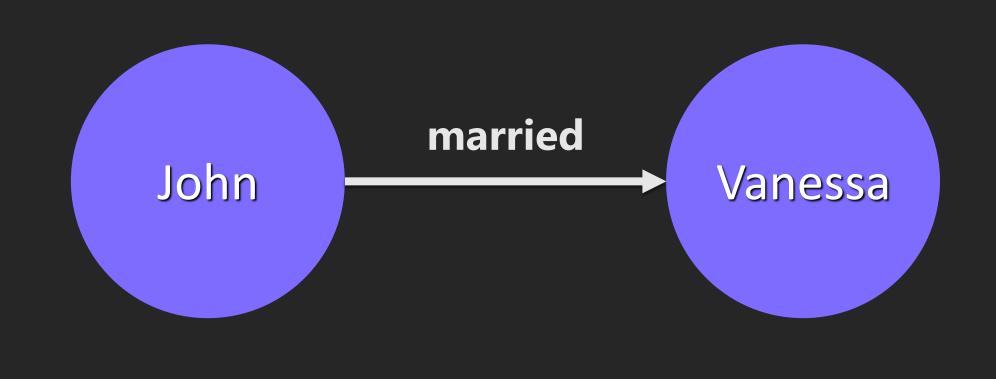
Hypergraphs generalize the common notion of graphs by relaxing the definition of edges

**Graph:** Edge = pair of vertices

**Hypergraph:** Hyperedge = set of vertices

### <u>Scenario – Traditional Marriage</u>

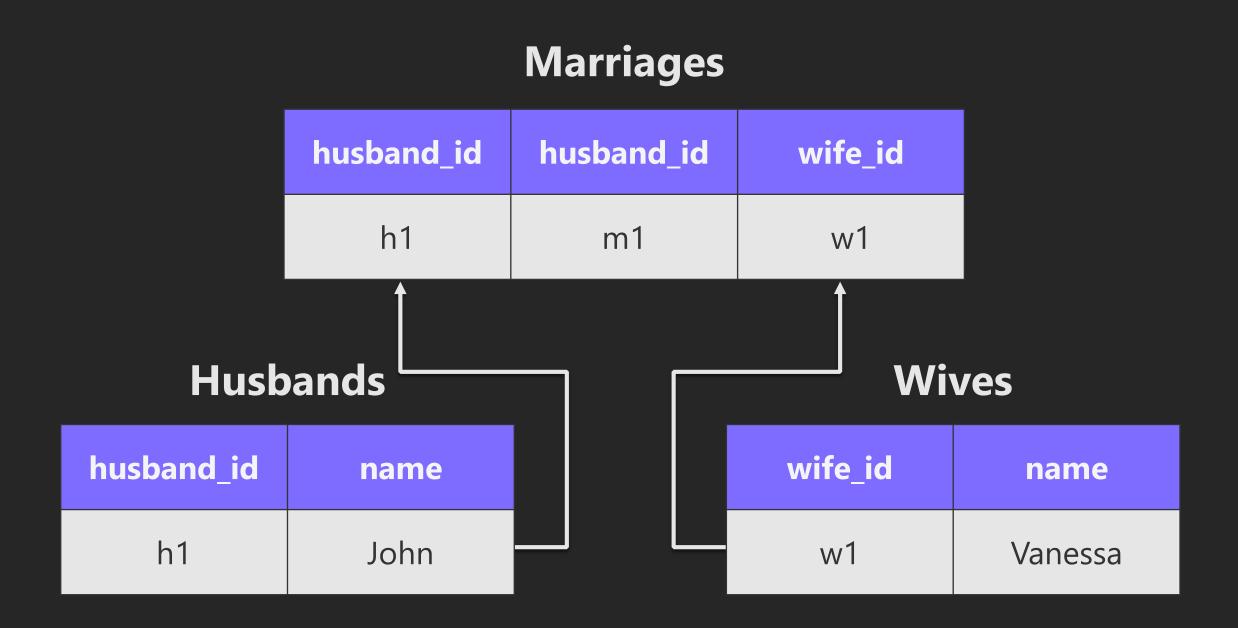
#### Marriages husband\_id husband\_id wife\_id h1 w1 m1Husbands Wives husband\_id wife\_id name name h1 John Vanessa w1

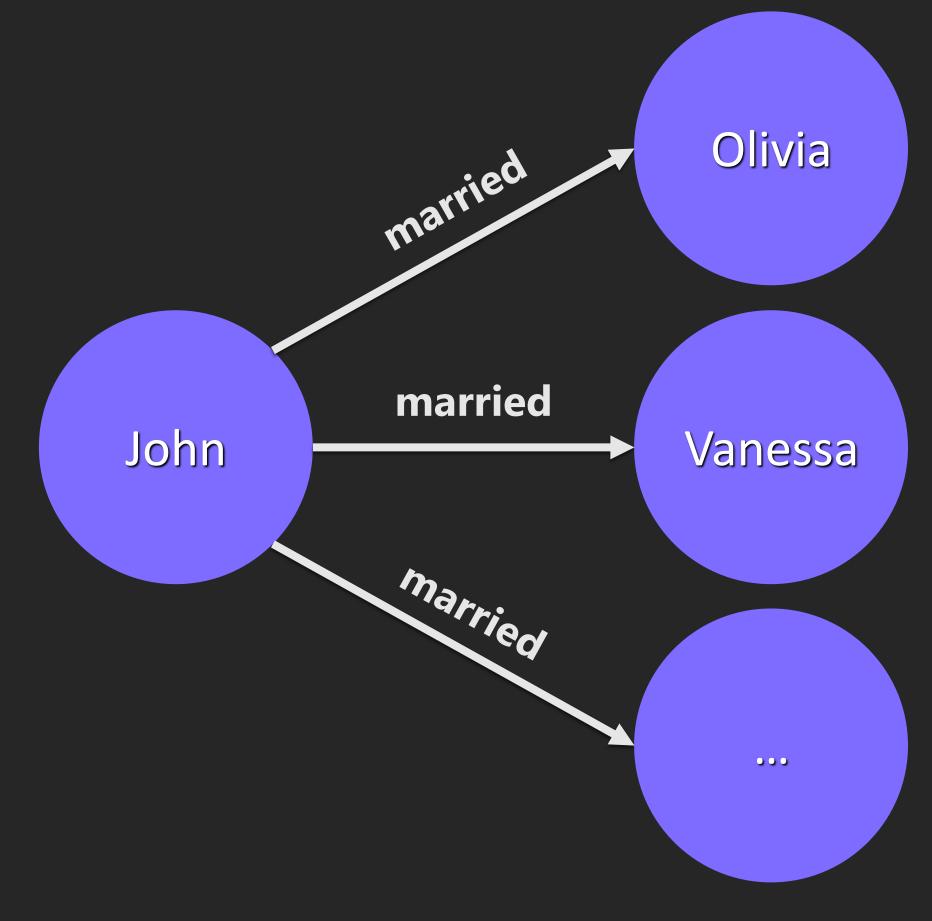


Relational

Directed

### Scenario – Polygamous Marriage

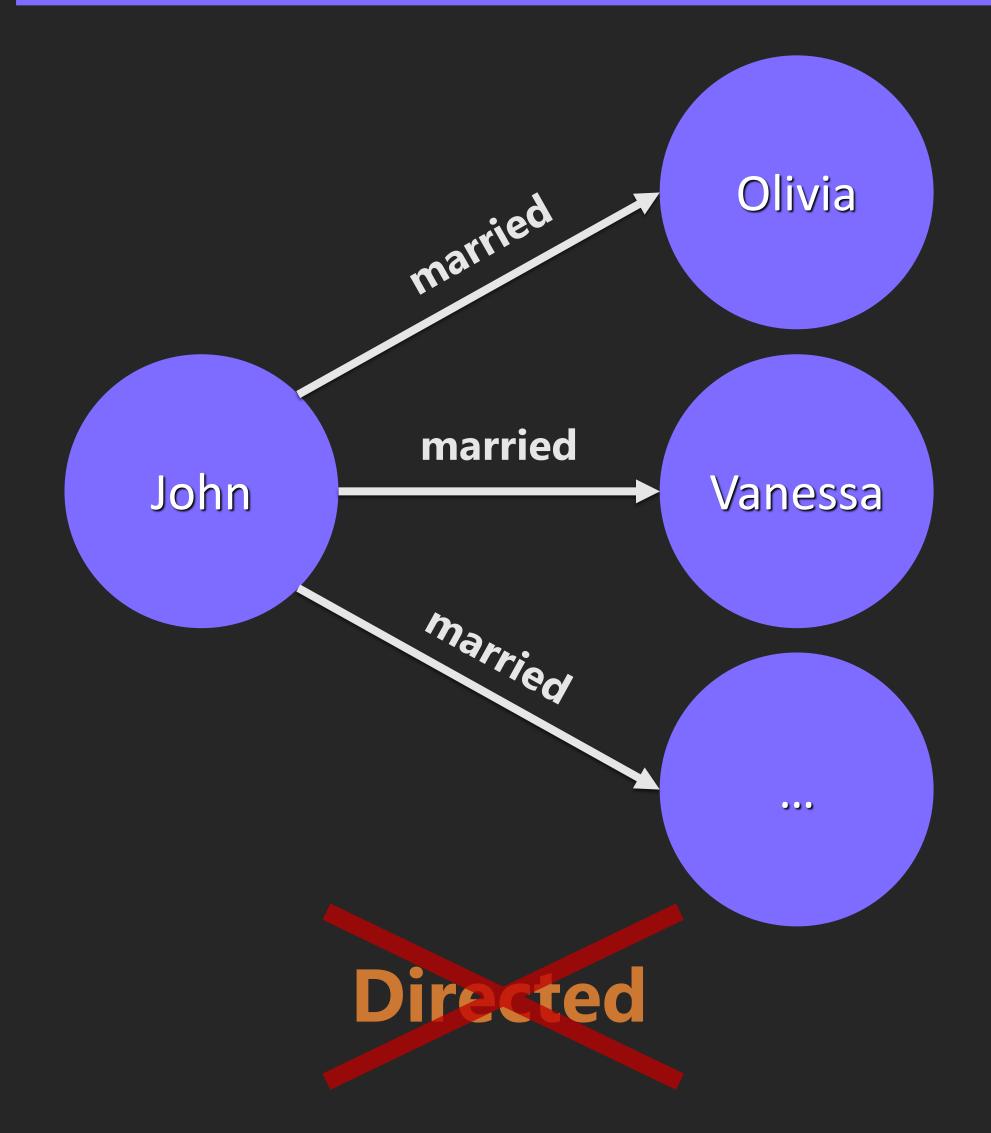


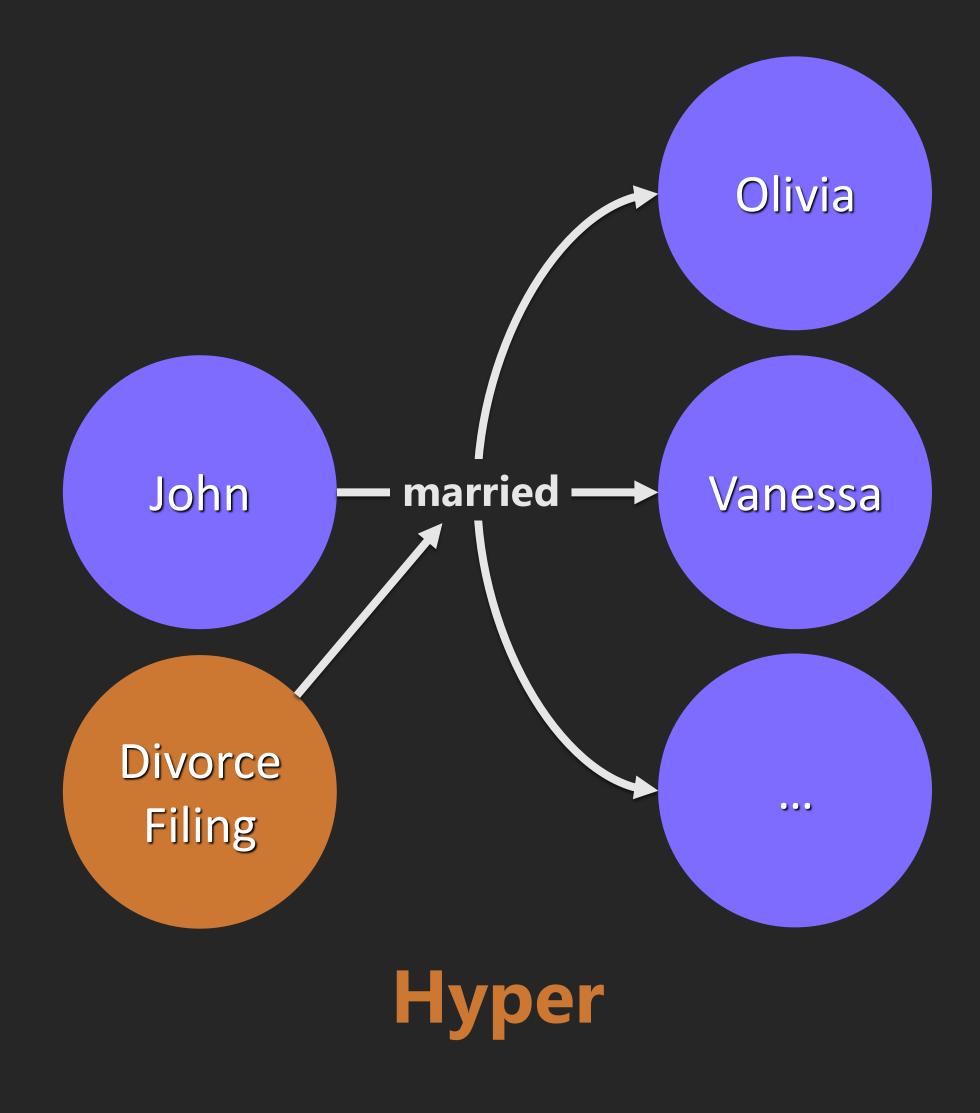




Directed

### Scenario - Divorce Filing





# How do I model a graph?

# Data Modeling

Technical Expertise

**Conceptual Model** 

Logical Model

Physical Model

### Conceptual Model

#### **Entities**

- Attribute value comprises a complex value type (e.g. address)
- Values with conceptual identities
- Value requires qualification via relation
- Example:
  - Find all recent orders delivered to the same delivery address (complex value type)

#### Relations

- Specify weight, strength, or some other quality about the relationship
- Example:
  - Find all my colleagues who are **level 2 or above** (relationship quality) in a **skill** (attribute value) we have in common

#### **Attributes**

- There's no need to qualify the relationship and consists of a simple value type (e.g. color)
- Have no conceptual identity (metadata)
- Example:
  - Find those projects written by contributors to my projects that use the same language (attribute value) as my projects

### Determine Entities

### Determine Entities

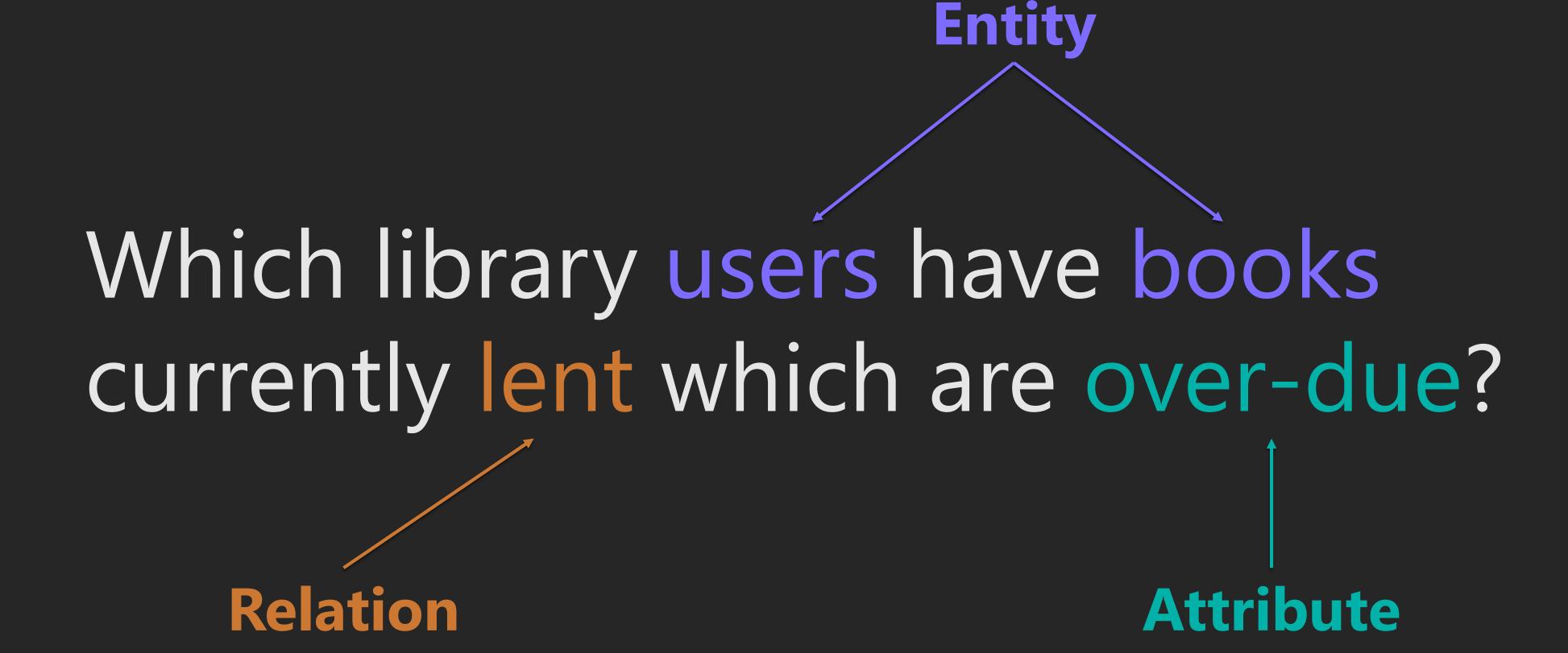
## Determine Relations

## Determine Relations

### Determine Attributes

### Determine Attributes

### Conceptual Model



### Entity-Relationship Model

Symbol Meaning Example Entity Employee Attribute Name Relationship Knows Link Employee Knows Name

### Draw Conceptual Model

### Objective

Which library users have books currently lent which are over-due?

### Data Available

#### Book

- Author
- Title
- Publish date
- Lend date
- Due date

#### User

- First/middle/last name
- DOB
- Email
- Gender

# Draw Conceptual Model

#### Objective

Which library users have books currently lent which are over-due?

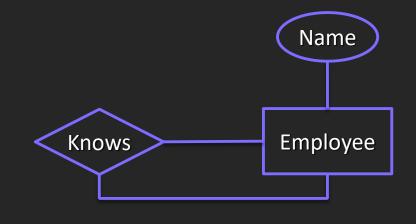
#### **Data Available**

#### Book

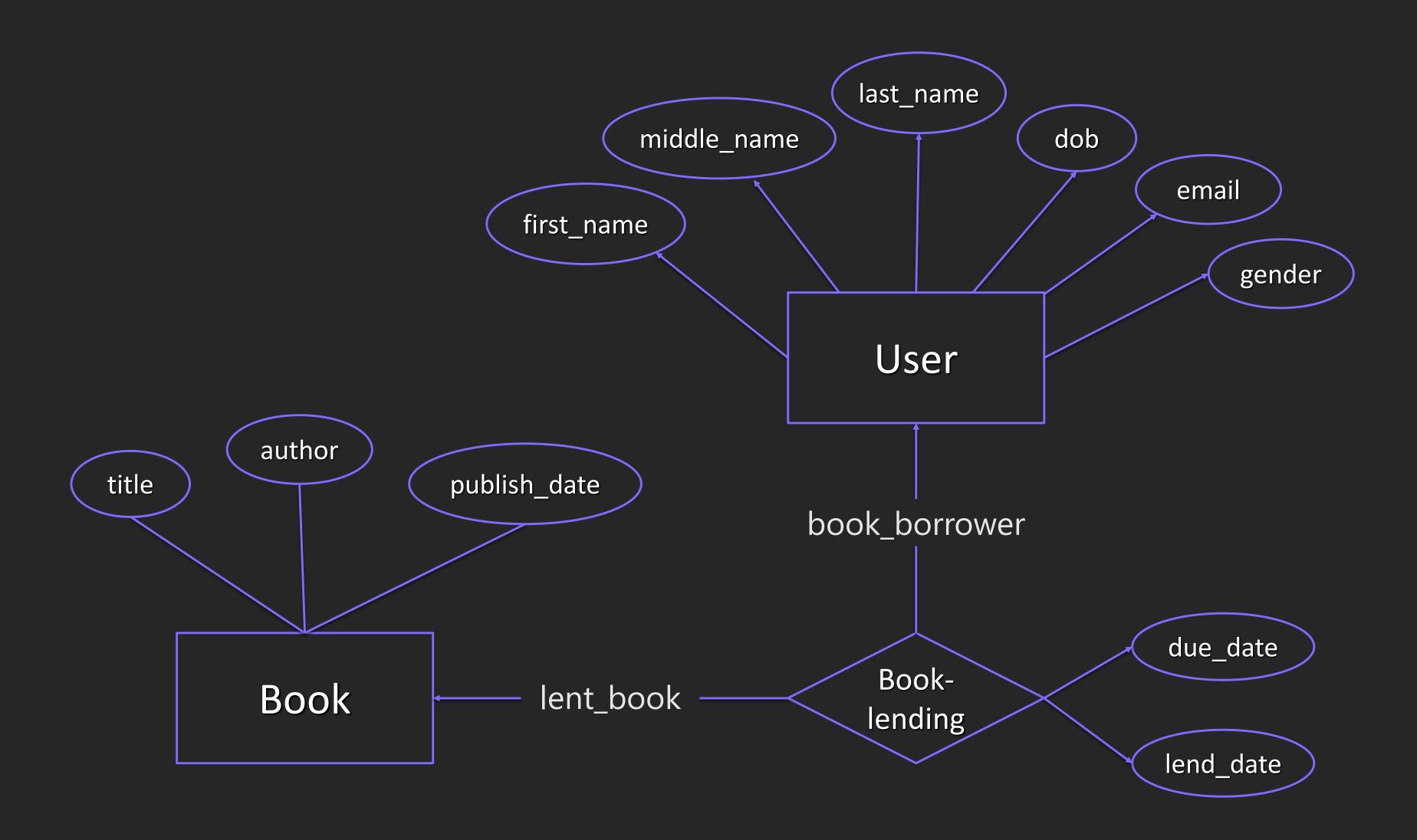
- Author
- Title
- Publish date
- Lend date
- Due date

#### User

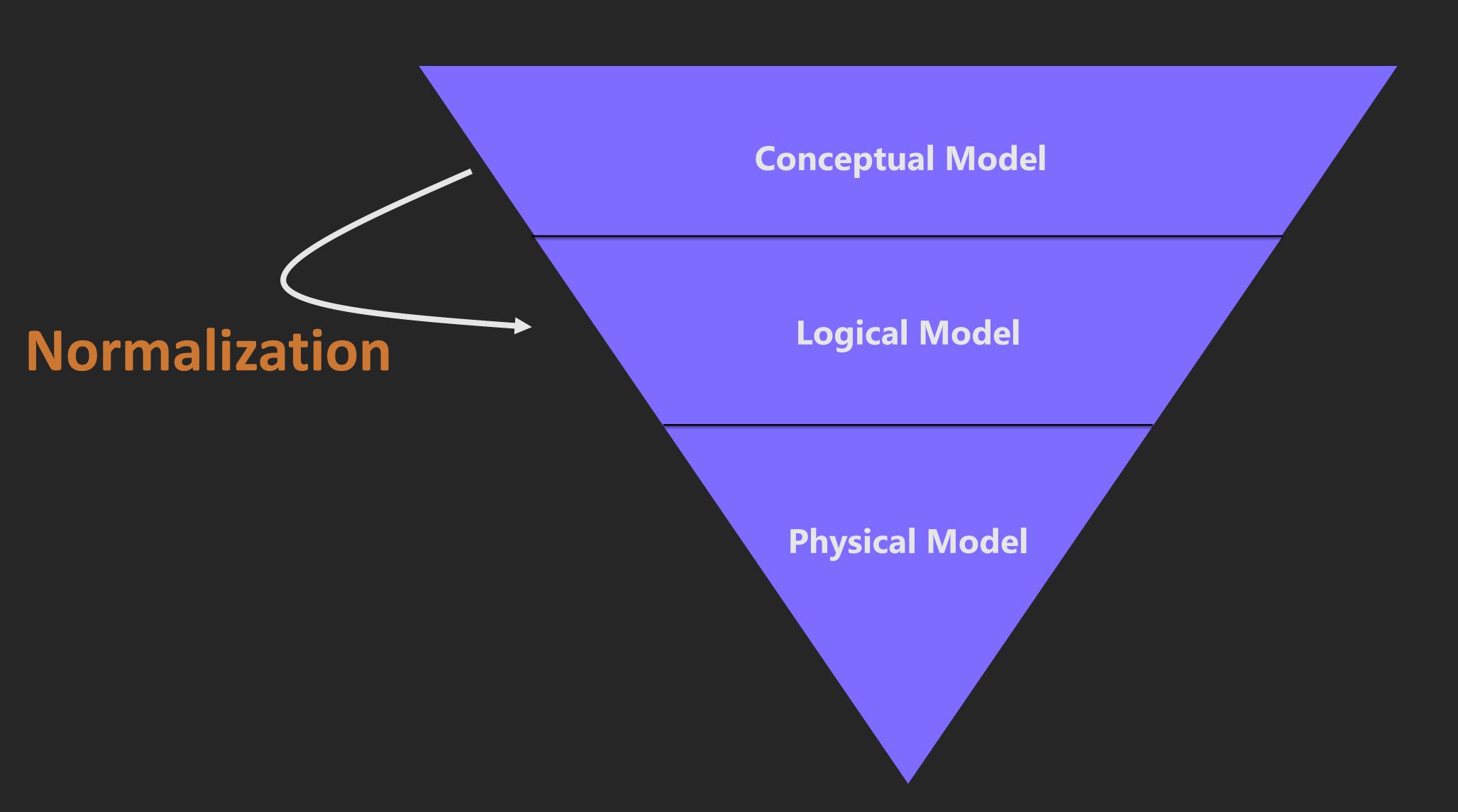
- First/middle/last name
- DOB
- Email
- Gender



# Conceptual Model



**Conceptual Model** Technical Expertise **Logical Model** 



### Users

user\_id

first\_name

middle\_name

last\_name

dob

email

gender

### **BookLendings**

user\_id

book\_id

lend\_date

due\_date

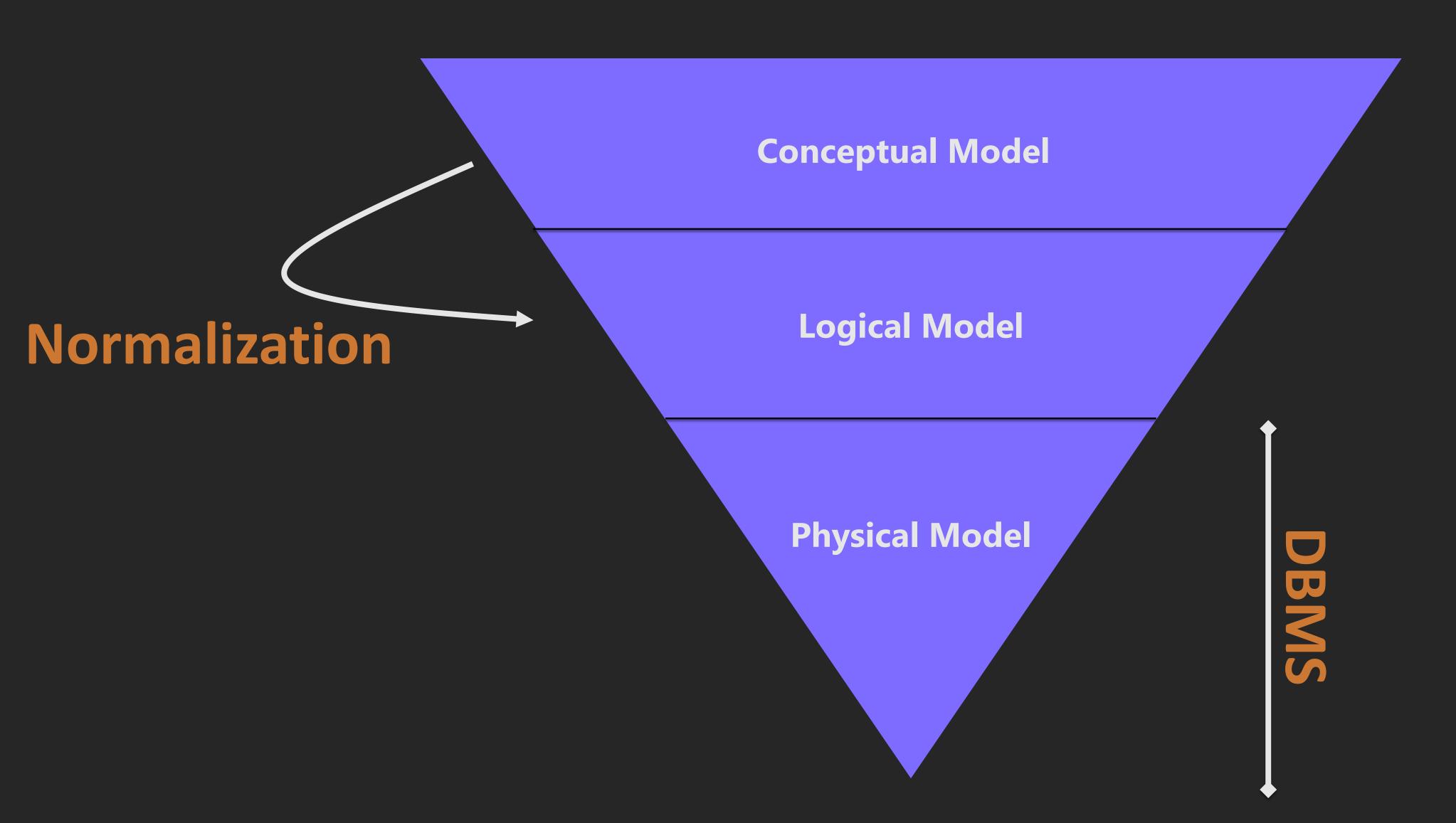
### Books

book\_id

title

author

publish\_date



# Introducing Grakn



# GRAKN.AI

Knowledge Representation System

### Users

user\_id

first\_name

middle\_name

last\_name

dob

email

gender

### **BookLendings**

user\_id

bookid

lend\_date

due\_date

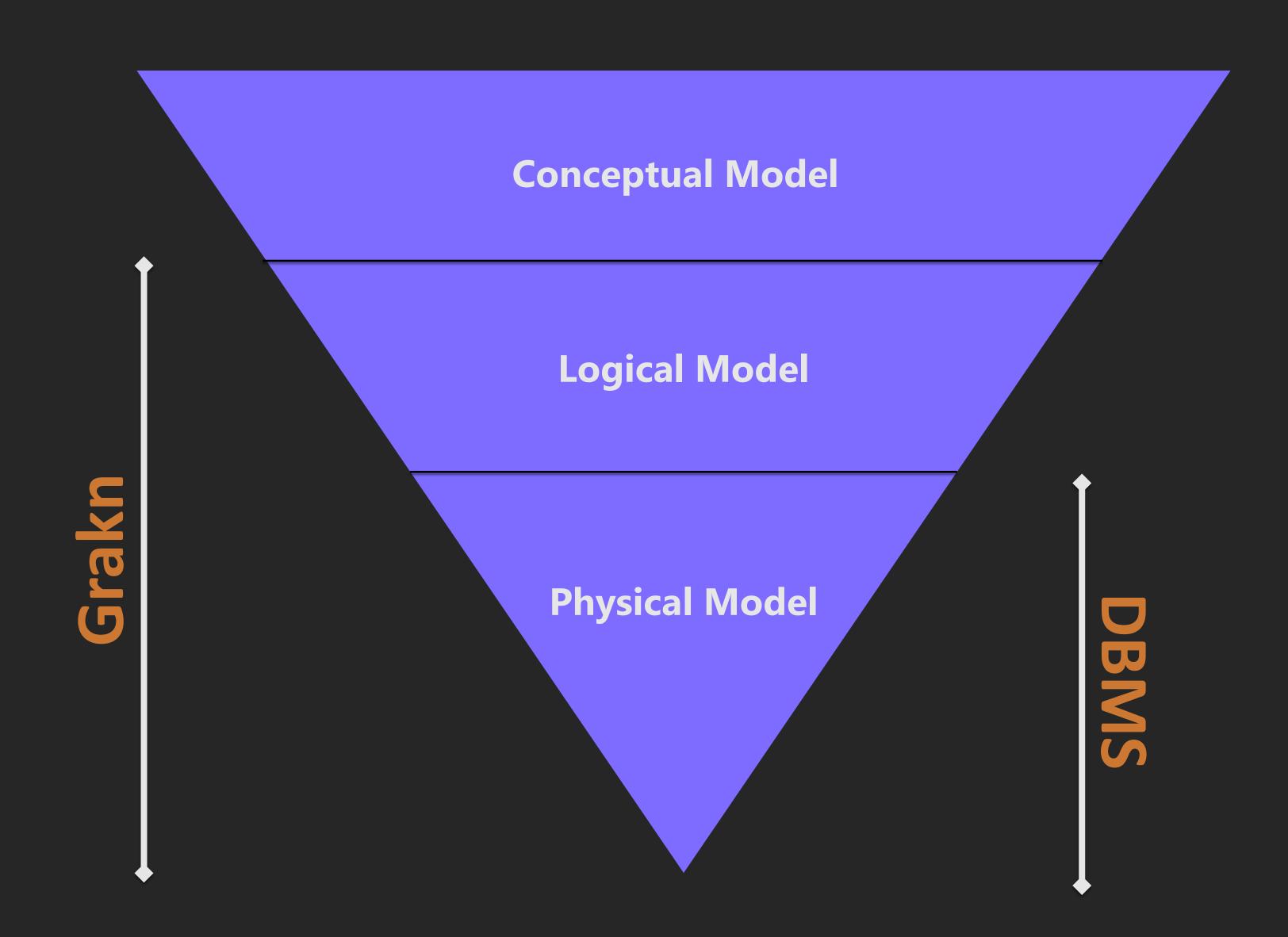
### Books

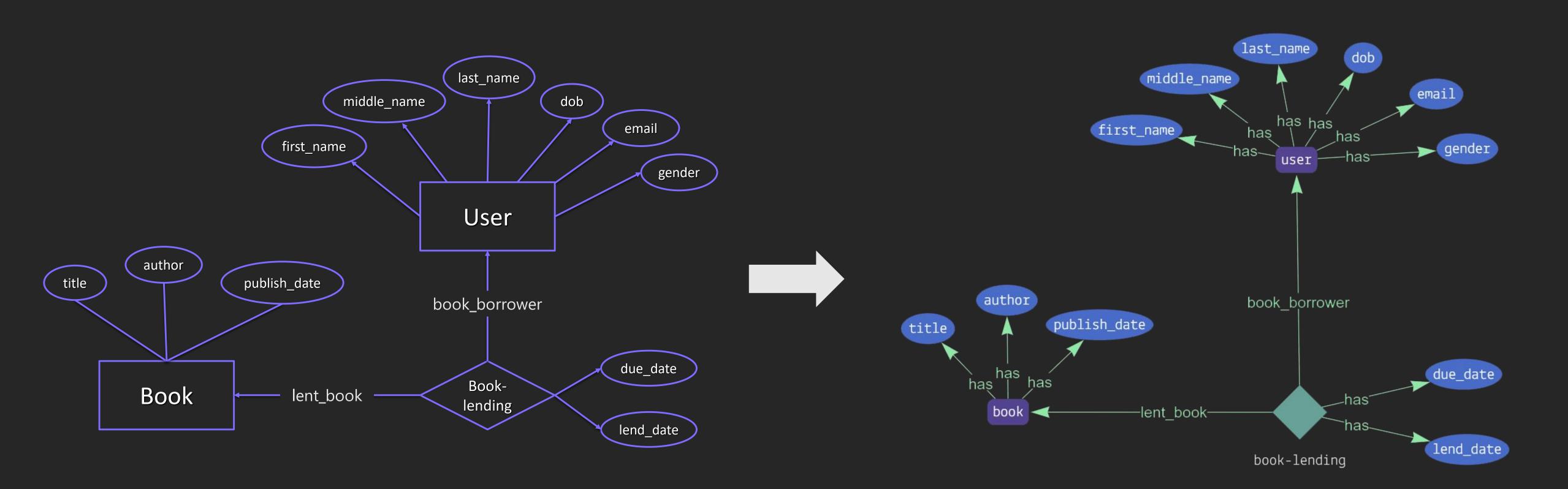
book\_id

title

author

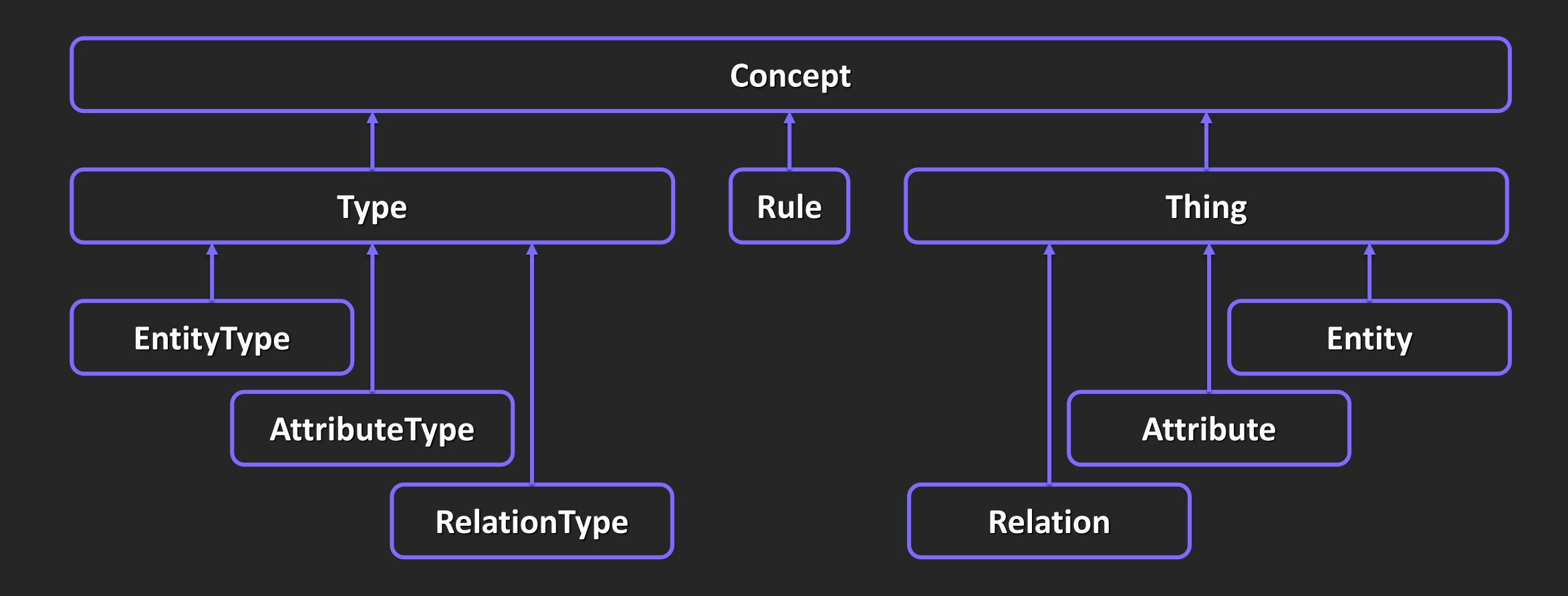
publish\_date





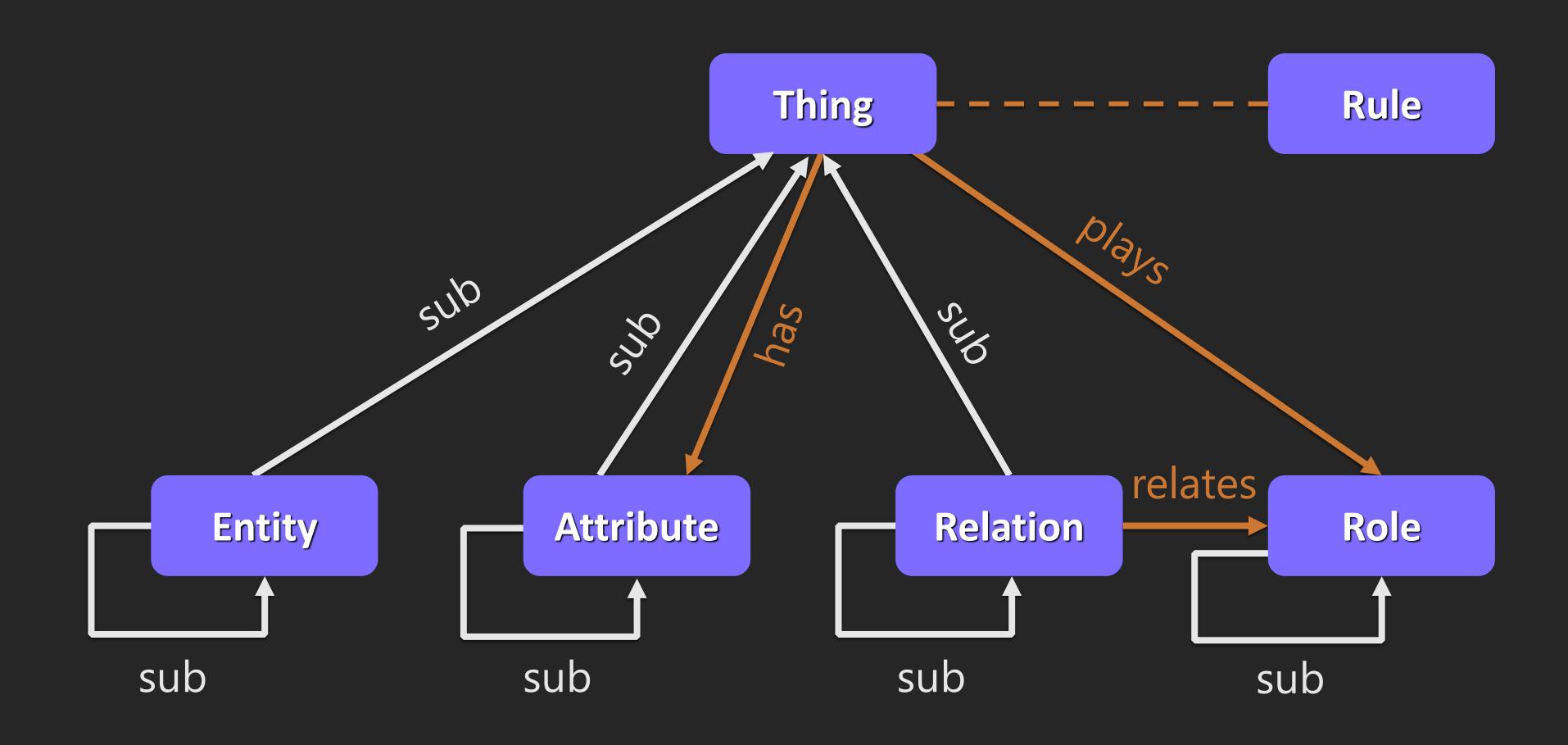
# Grakn Ontology

A highly expressive and intelligent type system for your complex data

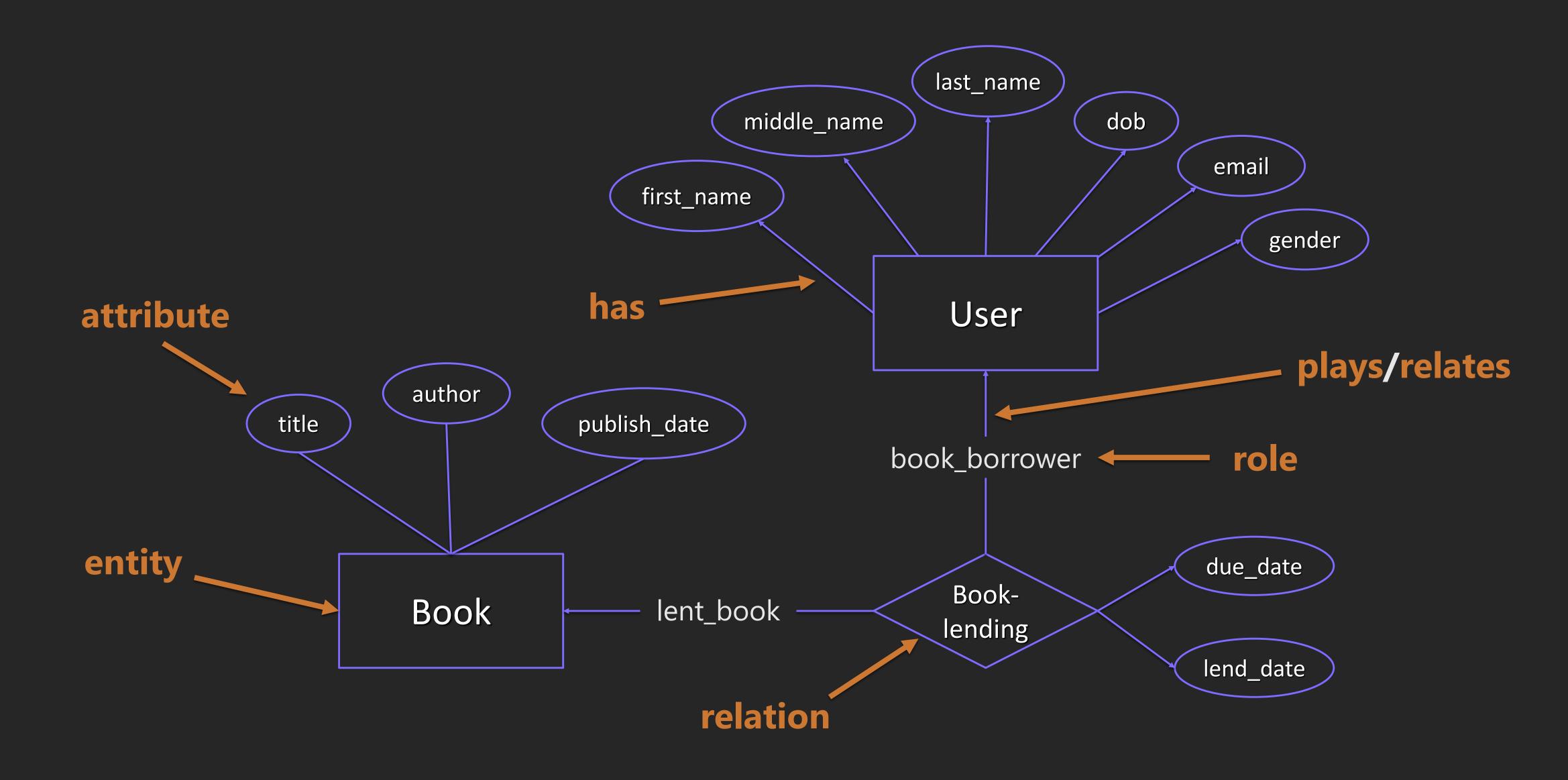


# Knowledge Model

A schema which can represent: type hierarchies, hyper-relations and rules



# Knowledge Model

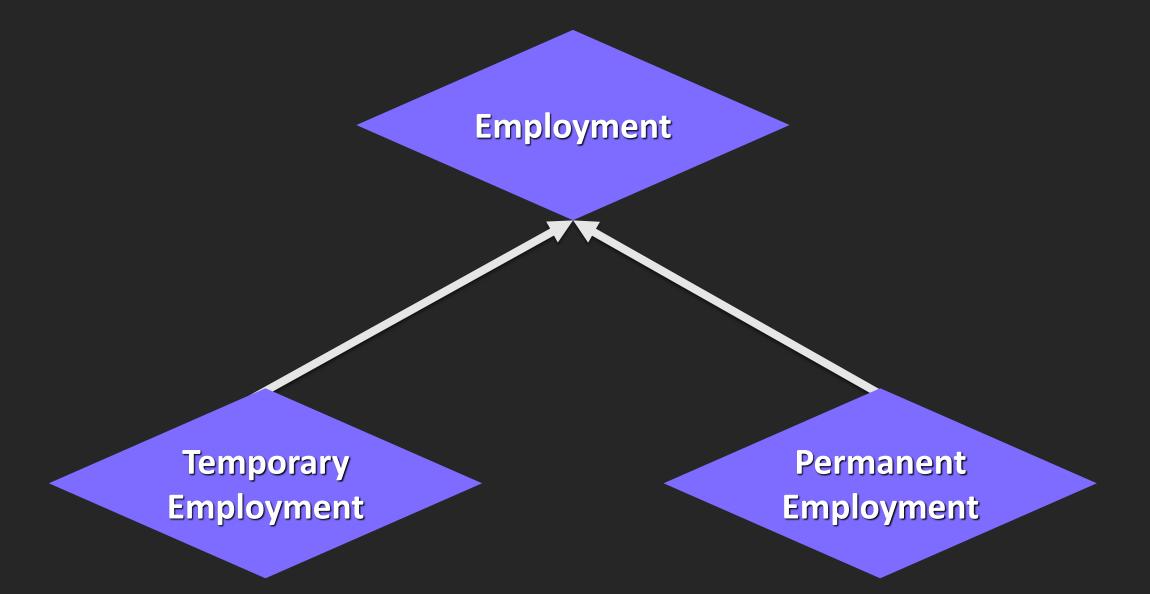


## <u>Grakn Modeling – Hierarchies</u>

### **Entity Type Hierarchies**

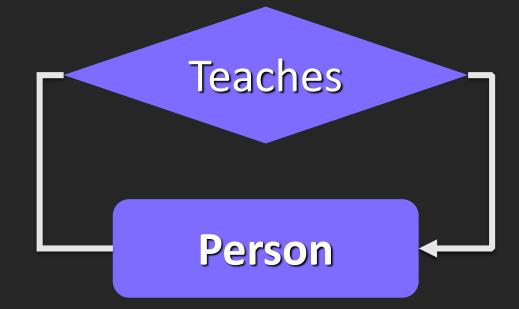
# Animal sub sub Mammal Reptile sub Sub Crocodile

### **Relation Type Hierarchies**

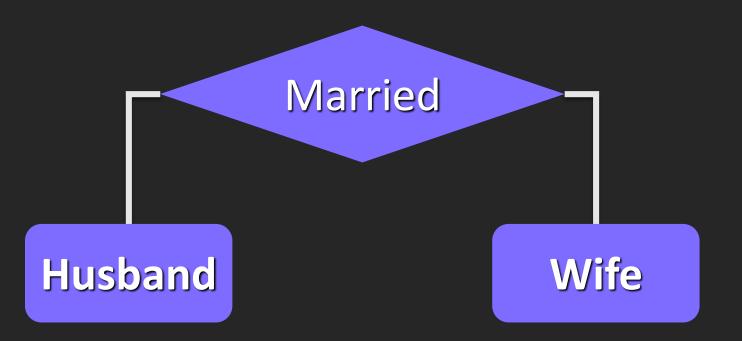


# Grakn Modeling – Relations

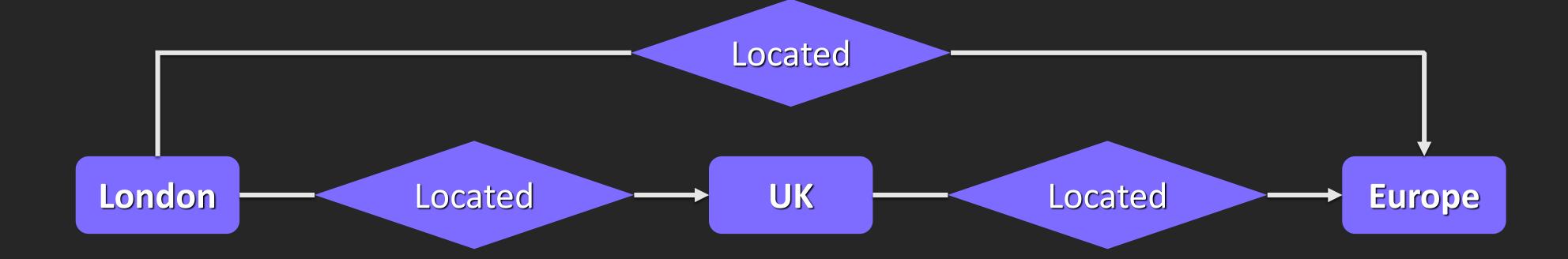
### **Reflexive Relations**



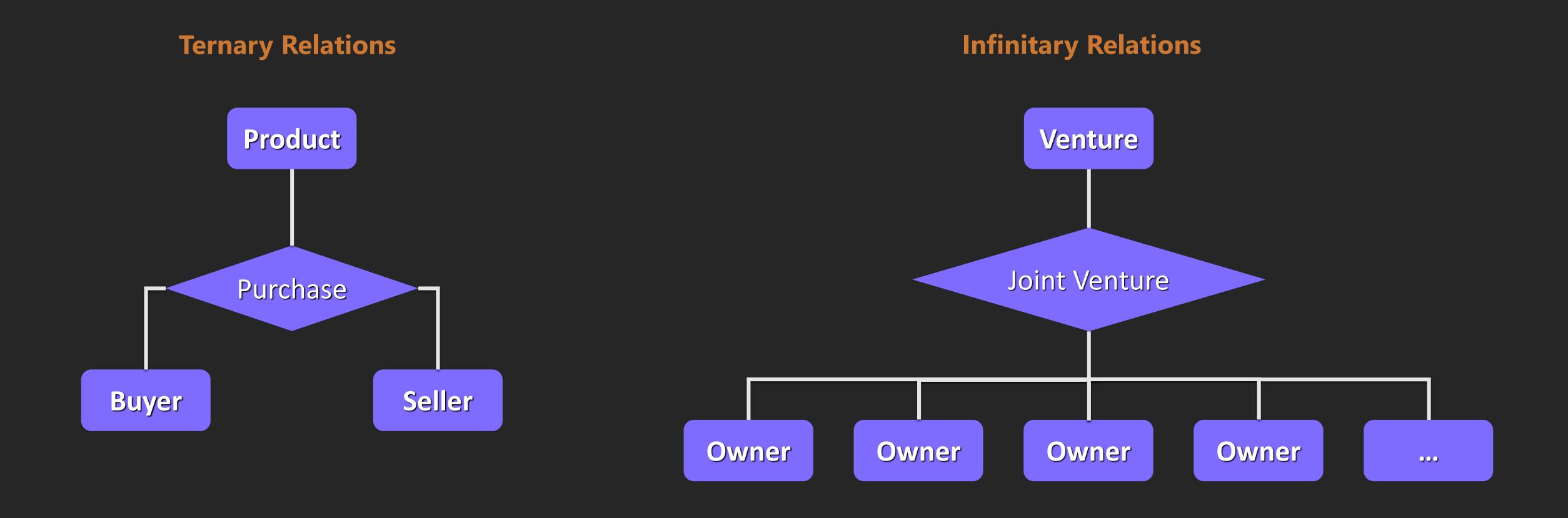
### **Symmetric Relations**



### **Transitive Relations**



# Grakn Modeling – Relations



# Grakn Modeling – Relations

# Nested Relations Located Married City Wife

# **Equivalent Relations** Parentship Parent Parent Child Child Siblingship

## Grakn Modeling – Tips

### Model incrementally

Start with questions you want answered now

Be as specific as possible

father > parent

Relations are first-class citizens

No foreign keys!

Generally: Tables are entities, rows are instances, columns are attributes

Associative/junction tables are relations

Unconnected data is hard to reach

Relations provide easy access to connected data

No such thing as nulls

The absence of value indicates the lack of value

Prefer ingesting non-changing data (facts) over syncing volatile data

Represent new data (new facts) with additional values

# Introducing Graq

# Commands

CREATE/ALTER TABLE	->	define
SELECT+WHERE	->	match+get
INSERT INTO	->	insert
DELETE	->	delete
DROP TABLE	->	undefine
UPDATE	->	match+delete+insert

# Data Types

NUMERIC -> long

DECIMAL -> double

TEXT -> string

BOOLEN -> boolean

DATE -> date

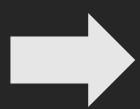
**Products** 

ProductId

ProductName

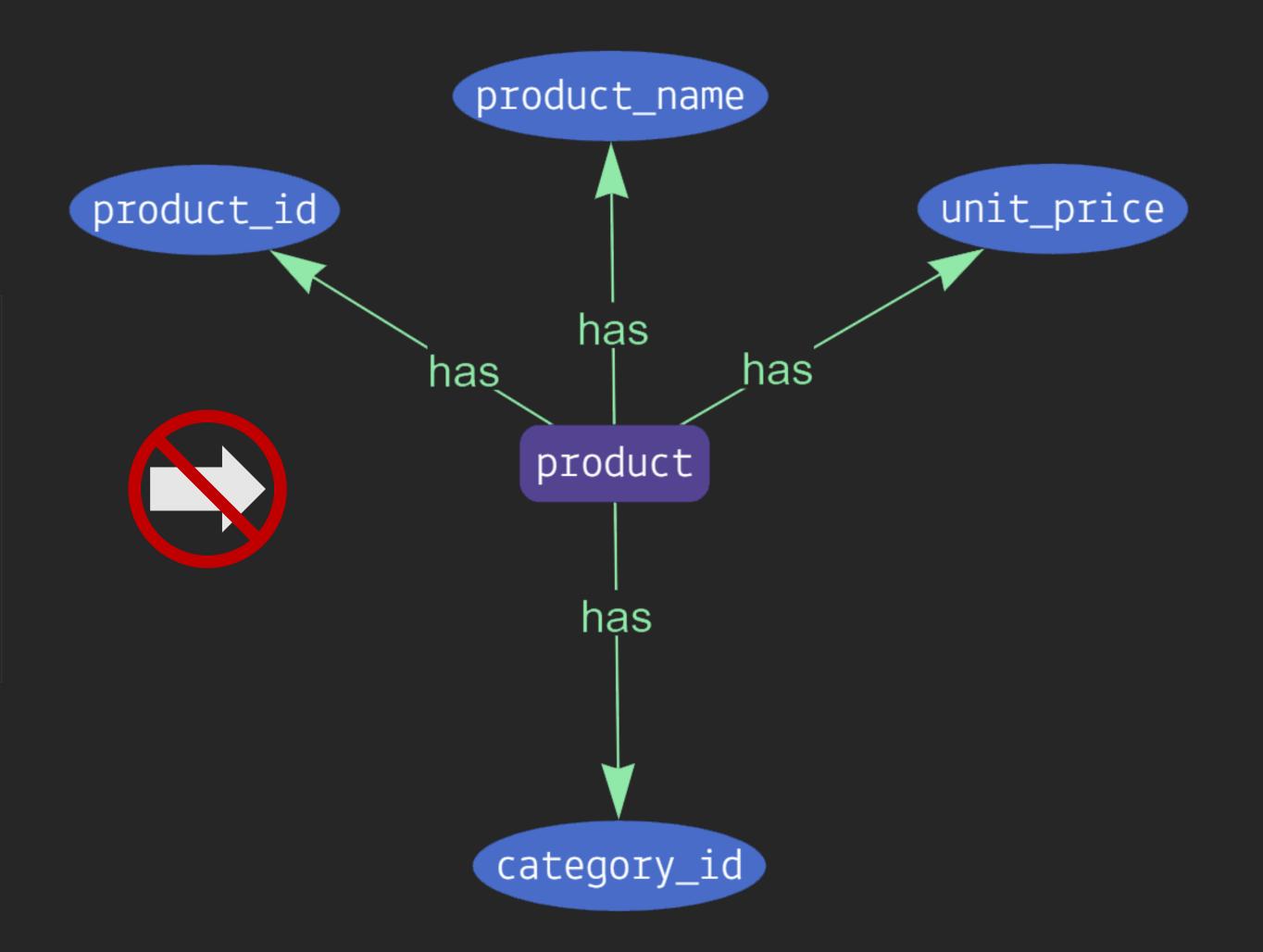
Categoryld

UnitPrice

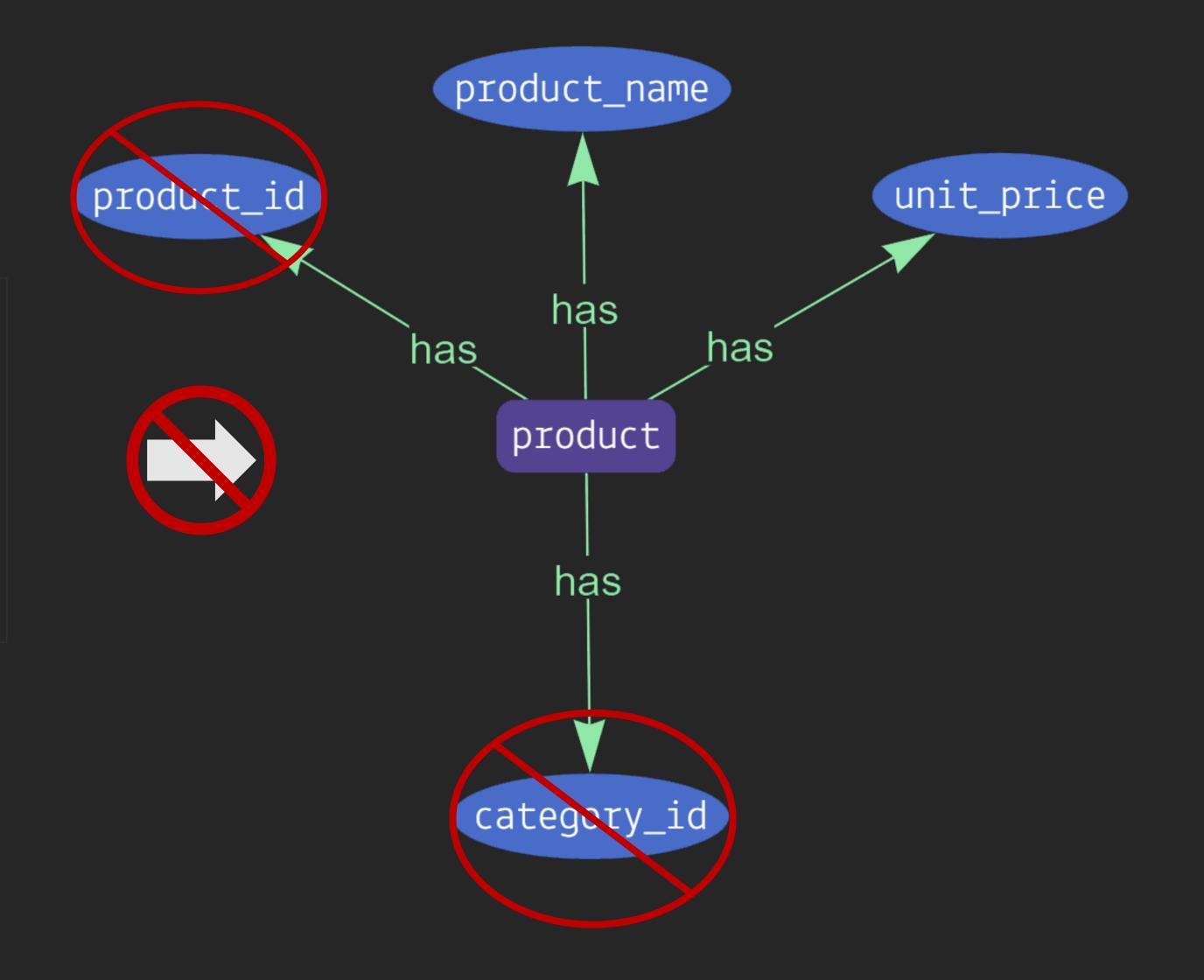


```
CREATE TABLE products (
    product_id smallint NOT NULL PRIMARY KEY,
    product_name character varying(40) NOT NULL,
    category_id smallint,
    unit_price float,
    FOREIGN KEY (category_id) REFERENCES categories
);
```

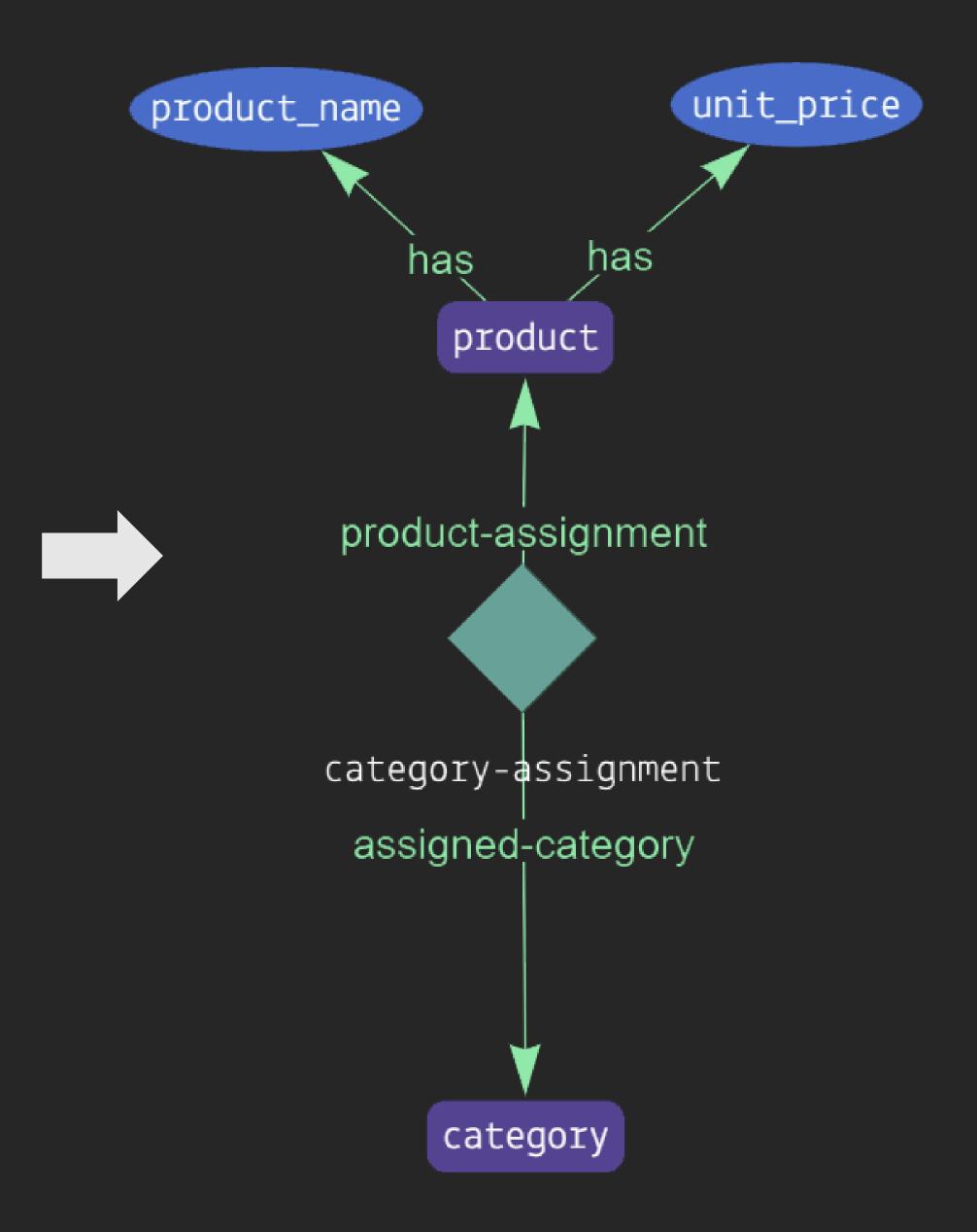
```
CREATE TABLE products (
    product_id smallint NOT NULL PRIMARY KEY,
    product_name character varying(40) NOT NULL,
    category_id smallint,
    unit_price float,
    FOREIGN KEY (category_id) REFERENCES categories
);
```



```
CREATE TABLE products (
    product_id smallint NOT NULL PRIMARY KEY,
    product_name character varying(40) NOT NULL,
    category_id smallint,
    unit_price float,
    FOREIGN KEY (category_id) REFERENCES categories
);
```



```
CREATE TABLE products (
    product_id smallint NOT NULL PRIMARY KEY,
    product_name character varying(40) NOT NULL,
    category_id smallint,
    unit_price float,
    FOREIGN KEY (category_id) REFERENCES categories
);
```



```
SQL
```

### Graql

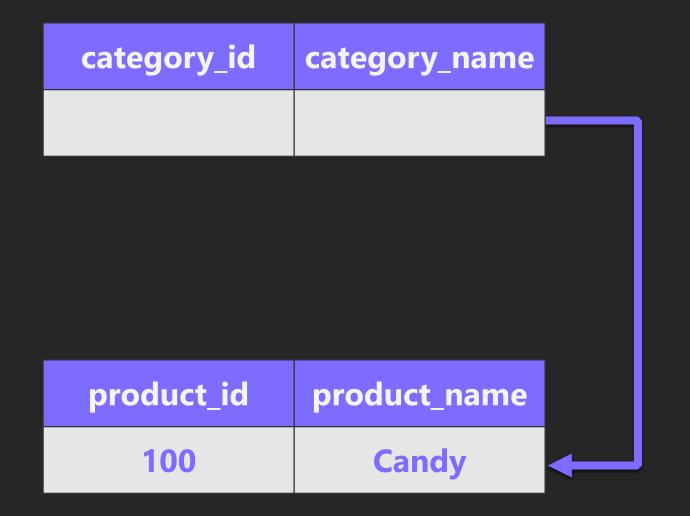
```
CREATE TABLE products (
  product_id smallint NOT NULL PRIMARY KEY,
  product_name character varying(40) NOT NULL,
  category_id smallint,
  unit_price float,
  FOREIGN KEY (category_id) REFERENCES categories
CREATE TABLE categories (
  category_id smallint NOT NULL PRIMARY KEY,
  category_name character varying(40)
```

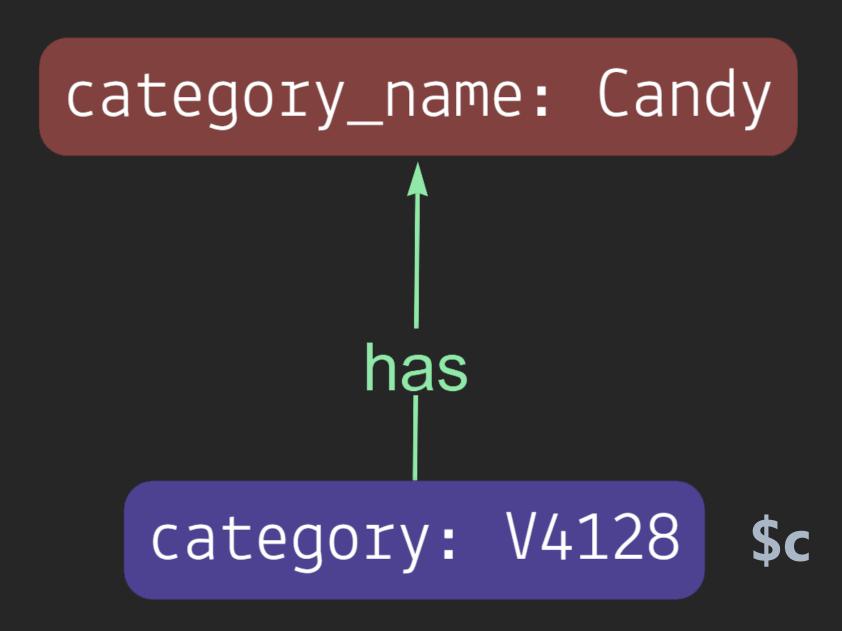
```
define
product sub entity,
  key product_name,
  has unit_price,
  plays product-assignment;
category sub entity,
  has category_name,
  plays assigned-category;
product_name sub attribute, datatype string;
unit_price sub attribute, datatype double;
category_name sub attribute, datatype string;
category-assignment sub relation,
  relates assigned-category,
  relates product-assignment;
```

```
★ Fill in the Blanks ★
```

```
CREATE TABLE albums (
  album id SMALLINT NOT NULL PRIMARY KEY,
  album_name CHARACTER VARYING(40) NOT NULL,
  release date DATE NOT NULL,
  artist id SMALLINT NOT NULL,
  FOREIGN KEY (artist id) REFERENCES artists
CREATE TABLE artists (
  artist_id smallint NOT NULL PRIMARY KEY,
  artist_name character varying(40)
```

```
album sub
  key album_name,
  has release_date,
        released album;
artist sub entity,
      artist_name,
  plays releasing artist;
album-release sub relation,
  relates released_album,
  relates releasing_artist;
album_name sub _____, datatype string;
release_date sub attribute, datatype date;
artist_name sub attribute, datatype
                     ★ Fill in the Blanks ★
```





```
INSERT INTO categories (category_id, category_name)
VALUES (100, 'Candy');
```

★ Fill in the Blanks ★

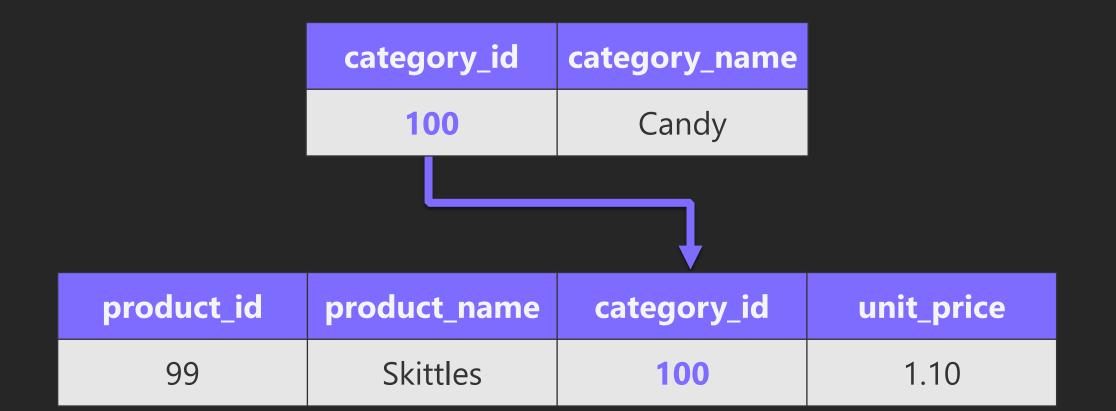
INSERT INTO artists (artist\_id, artist\_name)
VALUES (1, 'Michael Jackson');

\$a isa \_\_\_\_\_,
has \_\_\_\_\_ "Michael Jackson";

★ Fill in the Blanks ★

### SQL

### Graql



```
category_name: Candy

product_name: Skittles unit_price: 1.1

has has

category: V4128 $c product: V4240 $p

assigned-category product-assignment
```

```
INSERT INTO products
    (product_id, product_name, category_id, unit_price)
SELECT 99, 'Skittles', category_id, 1.10
FROM categories
WHERE category_name = 'Candy';
```

```
match
$c isa category, has category_name "Candy";
insert
$p isa product, has product_name "Skittles", has unit_price 1.10;
(product-assignment: $p, assigned-category: $c)
  isa category-assignment;
```

★ Fill in the Blanks ★

```
INSERT INTO albums
  (album_id, album_name, release_date, artist_id)
SELECT 1, 'Bad', to_date('1987-08-31','YYYYY-MM-DD'), artist_id
FROM artists
WHERE artist_name = 'Michael Jackson';
```

```
____ isa artist, ___ artist_name "Michael Jackson";
insert
___ isa album,
has album_name "Bad",
has release_date 1987-08-31;
(released_album: $al, releasing_artist: $ar) ___ album-release;
```

★ Fill in the Blanks ★

### SQL

### Graql

product_id	product_name	category_id	unit_price

```
$pn $up

product_name unit_price

has has

product
$p
```

SELECT product\_name, unit\_price
FROM products;

```
match
$p isa product,
  has product_name $pn,
  has unit_price $up;
get $pn, $up;
```

```
★ Fill in the Blanks ★
```

```
SELECT album_name, release_date FROM albums;
```

```
match
$a ___ album,
   has album_name $n,
   has release_date $r;
get ___, __;
```

★ Fill in the Blanks ★

product_id	product_name	category_id	unit_price
			>= 1.25
			>= 1.25
			>= 1.25

```
$pn $up >= 1.25

product_name unit_price

has has

product
$p
```

```
SELECT product_name, unit_price
FROM products
WHERE unit_price >= 1.25;
```

```
match
$p isa product,
   has product_name $pn,
   has unit_price $up;
$up >= 1.25;
get $pn, $up;
```

```
★ Fill in the Blanks ★
```

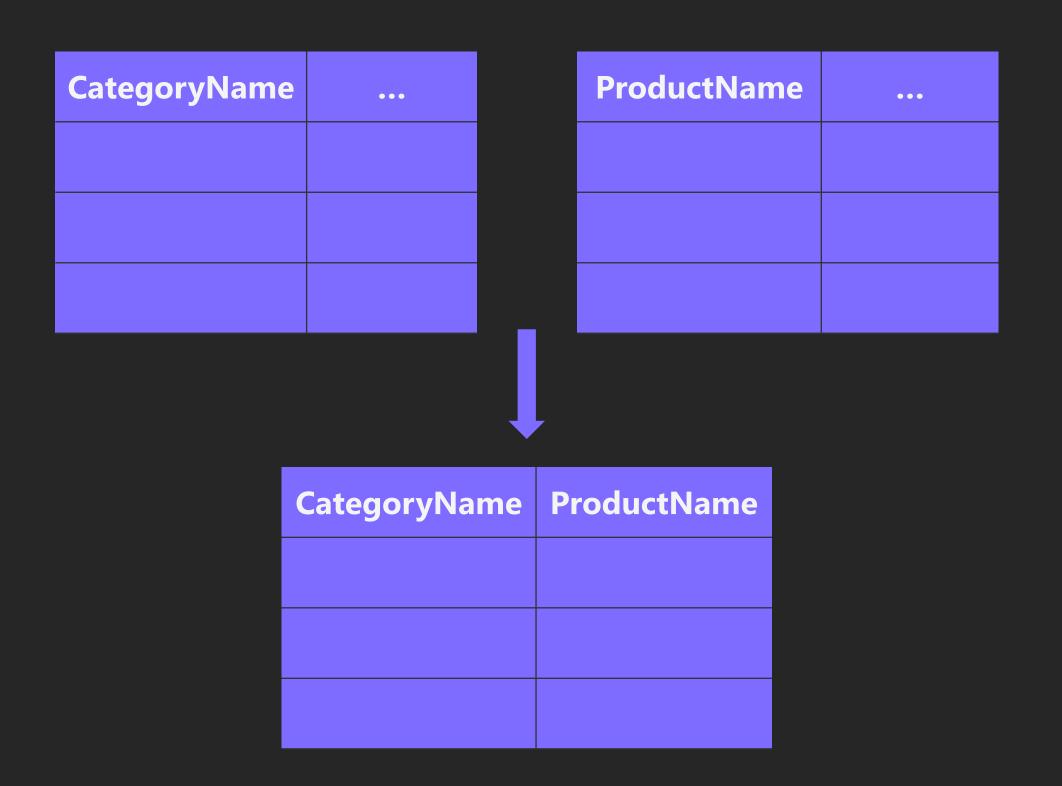
```
SELECT album_name, release_date
FROM albums
WHERE release_date < to_date('1990-01-01','YYYYY-MM-DD');</pre>
```

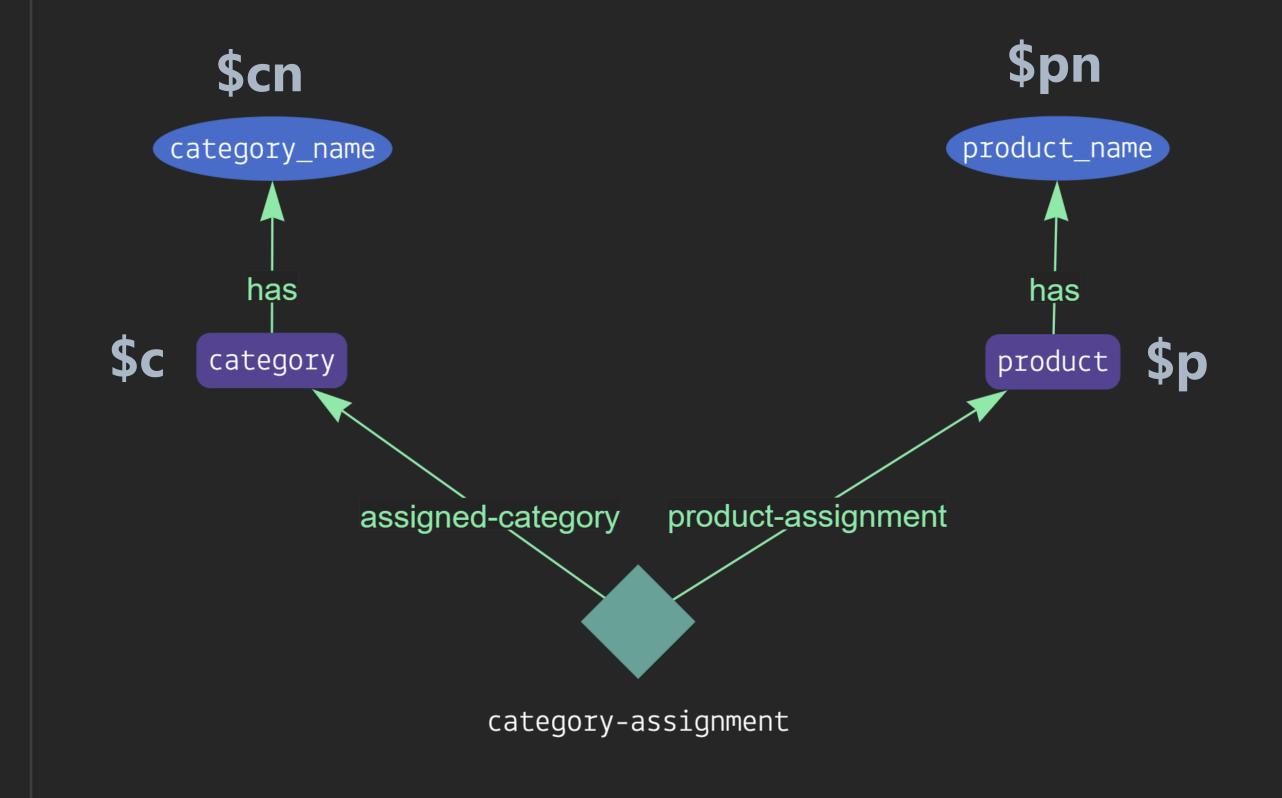
```
match
$a isa album,
has album_name $n,
has release_date $r;
$r < 1990-01-01;</pre>
```

★ Fill in the Blanks ★



### Graql





```
SELECT category_name, product_name
FROM categories
INNER JOIN products
ON categories.category_id = products.category_id;
```

### match

```
$c isa category, has category_name $cn;
$p isa product, has product_name $pn;
($c, $p) isa category-assignment;
get $cn, $pn;
```

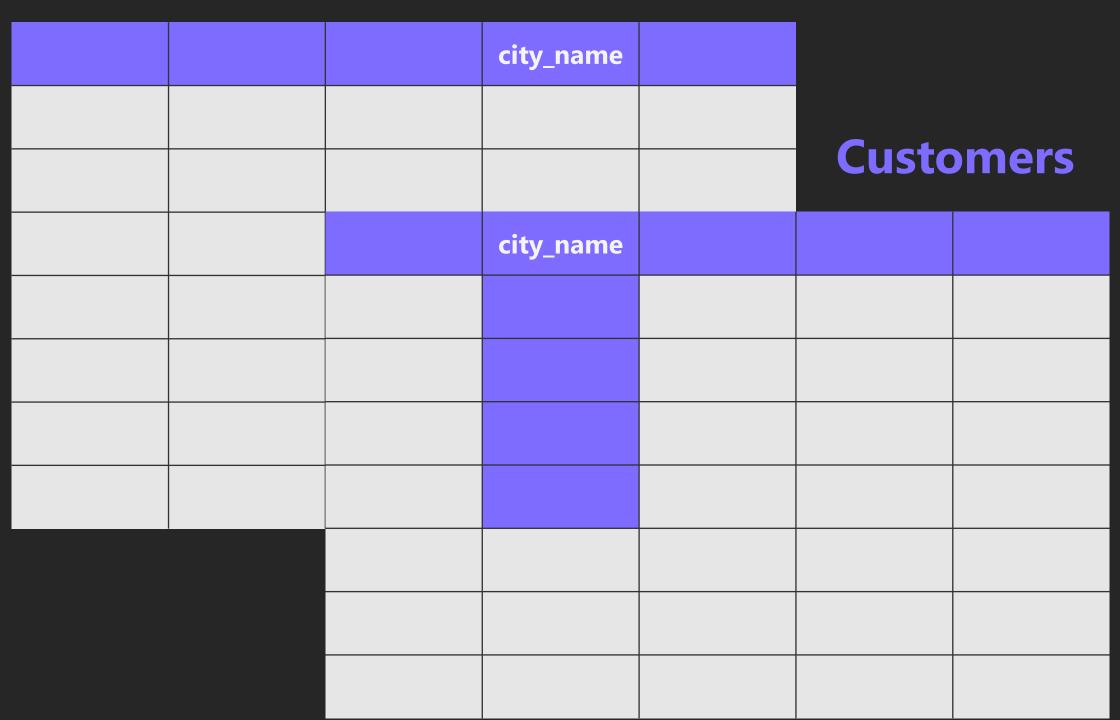
```
★ Fill in the Blanks ★
```

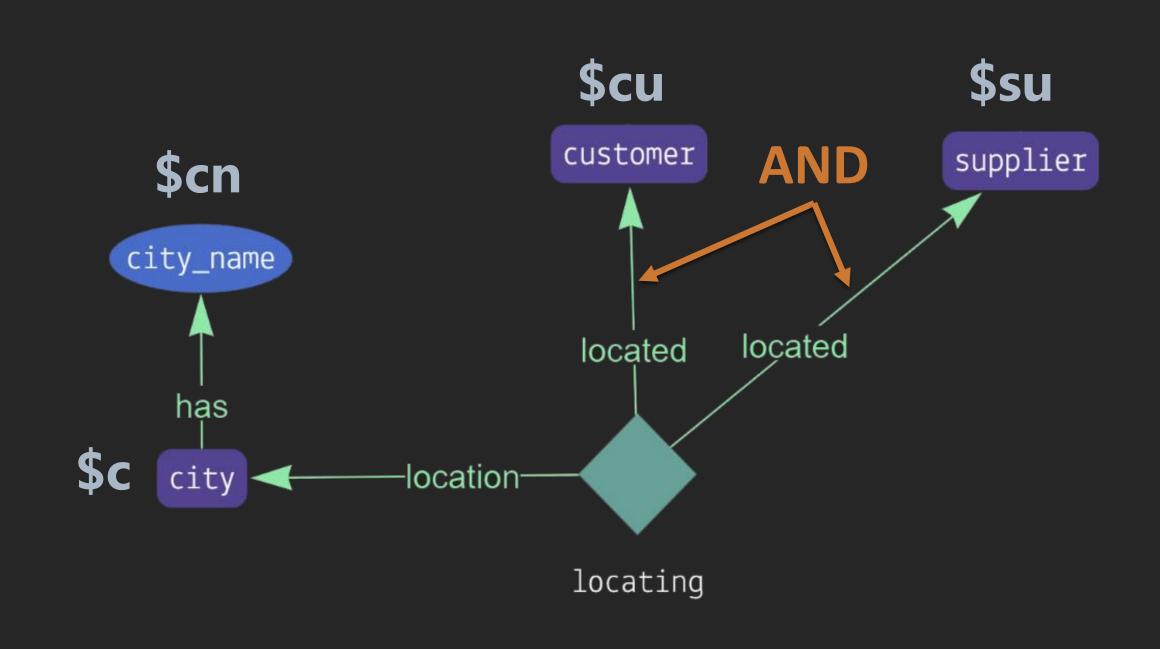
```
SELECT artist_name, album_name
FROM artists
INNER JOIN albums
ON artists.artist_id = albums.artist_id;
```

```
match
$ar isa artist,
  has artist_name $ar_name;
$al isa album,
  has album_name $al_name;
(___, ___) isa album-release;
get $ar_name, $al_name;
```

# Graql

### **Suppliers**





SELECT city\_name
FROM customers
INTERSECT
SELECT city\_name
FROM suppliers;

match

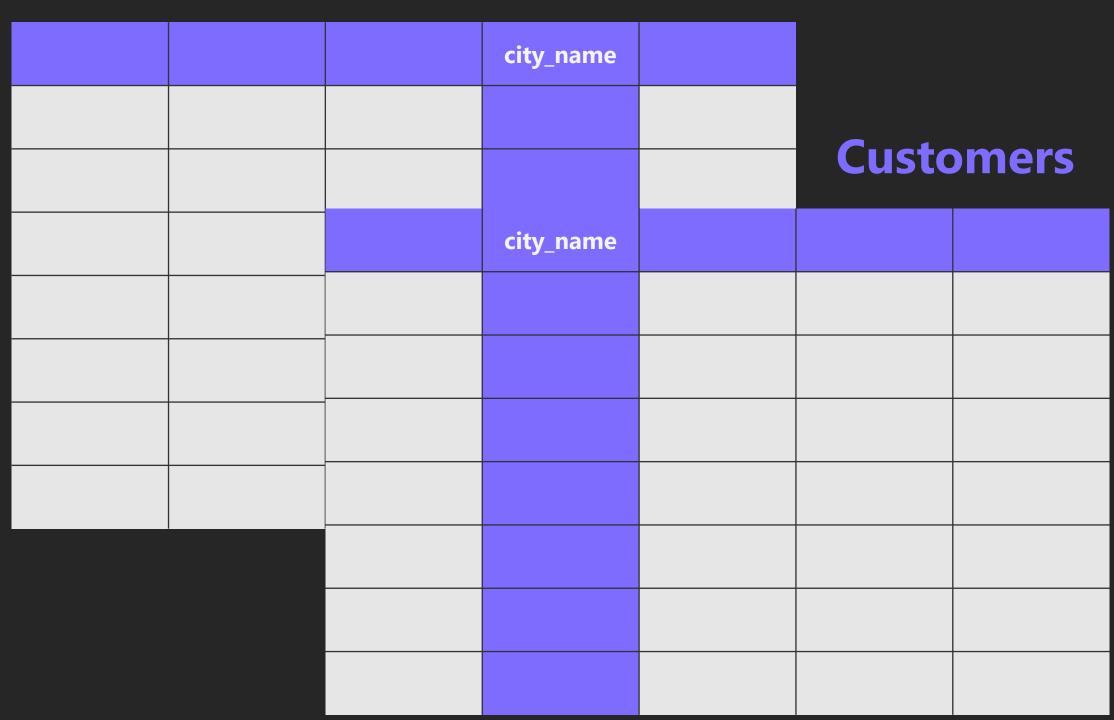
\$c isa city, has city\_name \$cn; \$cu isa customer; (location: \$c, \$cu); \$su isa supplier; (location: \$c, \$su); get \$cn;

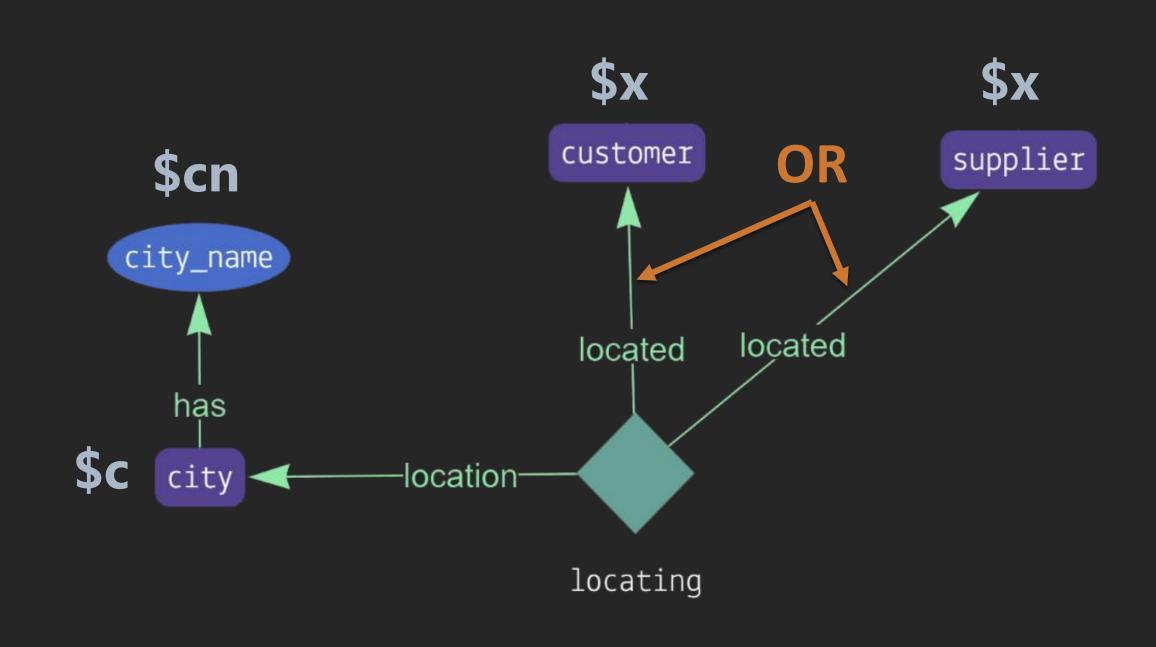
```
FROM albums
INNER JOIN artists on albums.artist_id = artists.artist_id
WHERE artist_name = 'Michael Jackson'
INTERSECT
SELECT album_name
FROM albums
WHERE release_date >= to_date('2000-01-01','YYYY-MM-DD');
```

```
$al ___ album, has album_name __;
$ar isa artist, ___ artist_name "Michael Jackson";
(released_album: ___, releasing_artist: $ar) isa album-release;
$al has release_date >= ____;
get $n;
```

# Graql

### **Suppliers**





SELECT city\_name
FROM customers
UNION
SELECT city\_name
FROM suppliers;

match
\$c isa city, has city\_name \$cn;
{\$x isa supplier;} or {\$x isa customer;};
(location: \$c, \$x);
get \$cn;

```
SQL
```

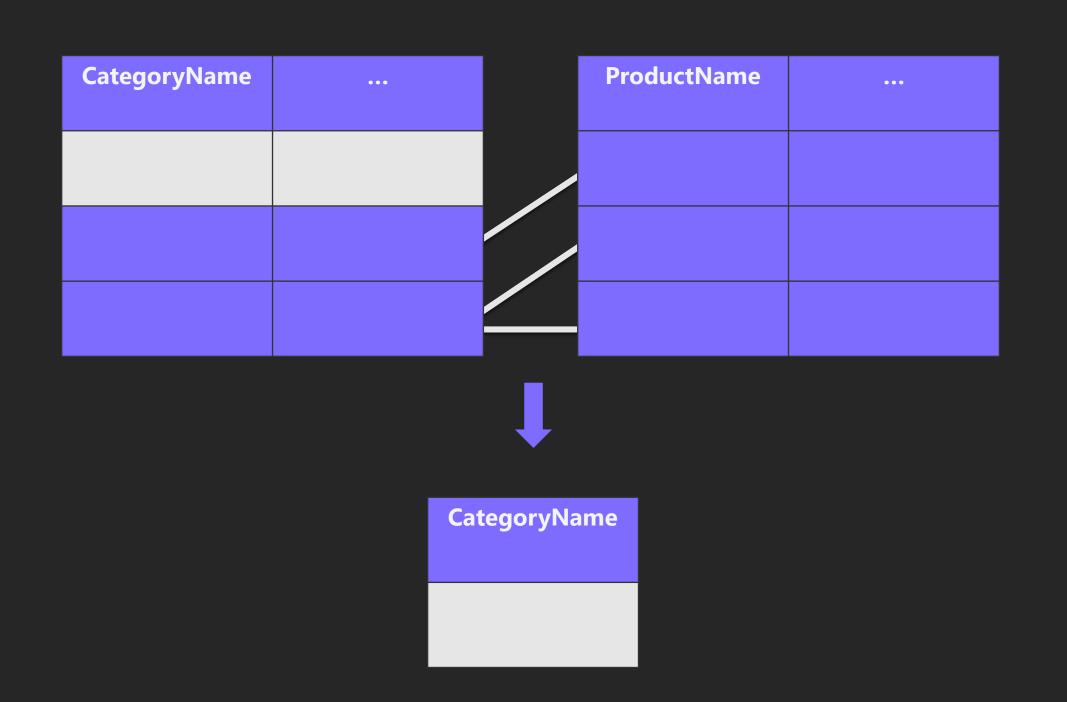
# Graql

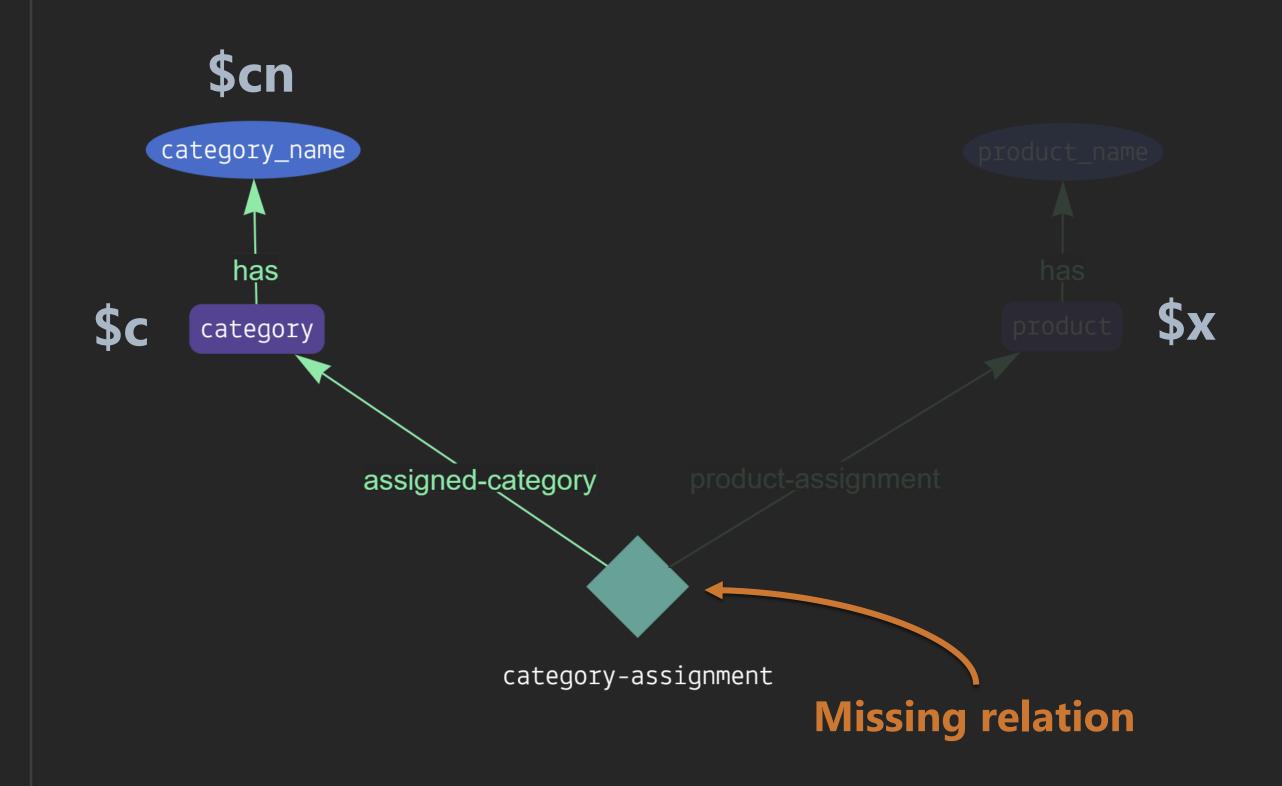
★ Fill in the Blanks ★

```
SELECT album_name
FROM albums
INNER JOIN artists on albums.artist_id = artists.artist_id
WHERE artist_name = 'Michael Jackson'
UNION
SELECT album_name
FROM albums
WHERE release_date >= to_date('2000-01-01','YYYYY-MM-DD');
```

```
match
_____, ____, $n;
{
    $ar isa artist, has artist_name "Michael Jackson";
    (released_album: $al, releasing_artist: $ar) isa album-release;
} ___ {
    $al has release_date >= 2000-01-01;
};
get $n;
```

# Graql





```
SELECT category_name
FROM categories
LEFT JOIN products
   ON categories.category_id = products.category_id
WHERE products.category_id IS NULL;
```

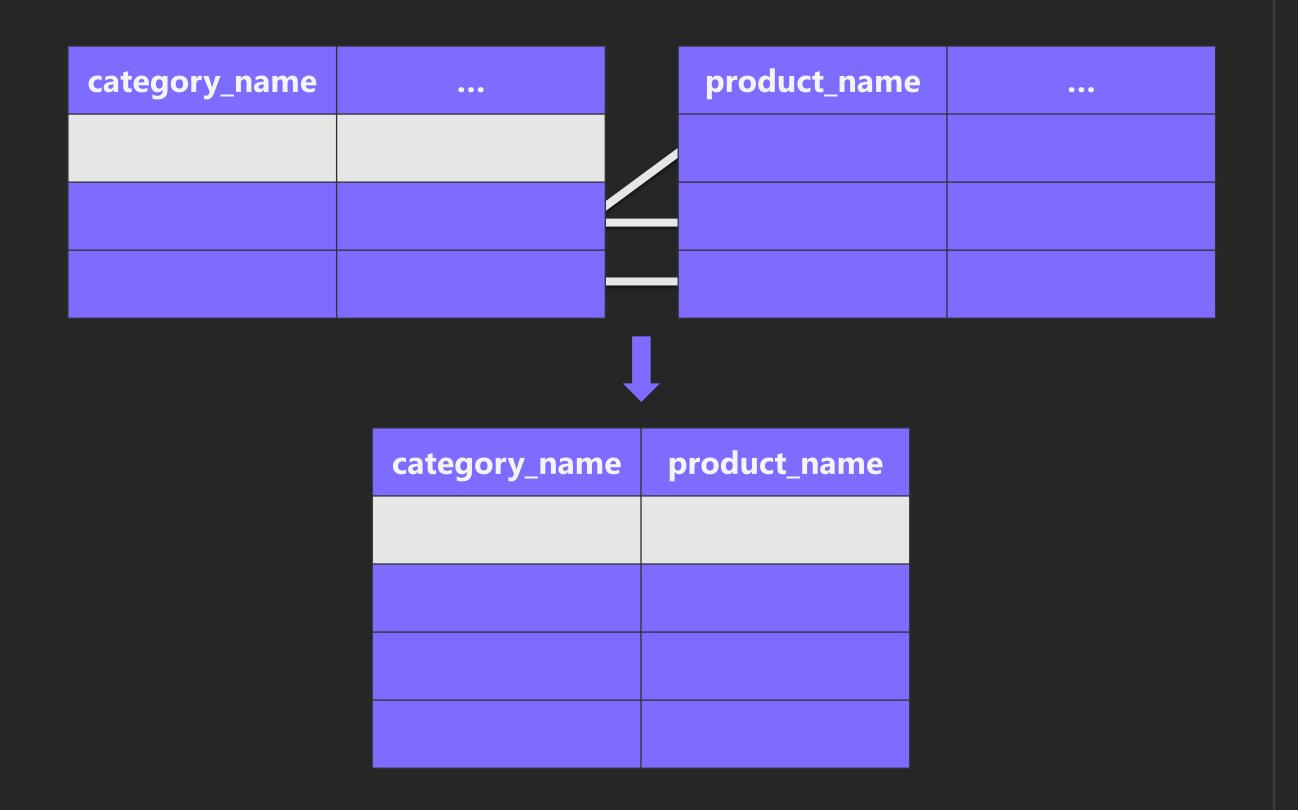
```
match
$c isa category,
  has category_name $cn;
not { ($c, $x) isa category-assignment; };
get $cn;
```

```
★ Fill in the Blanks ★
```

```
SELECT artist_name
FROM artists
LEFT JOIN albums
ON artists.artist_id = albums.artist_id
WHERE albums.artist_id IS NULL;
```

```
$ar isa artist,
has artist_name $ar_name;
{
   (___, $x) __ album-release;
};
___$ar_name;
```

# Graql



```
$cn

category_name

product_name

has

has

product $p

category-assignment

Optional relation
```

```
SELECT category_name
FROM categories
FULL OUTER JOIN products
ON categories.category_id = products.category_id;
```

### match

```
$c isa category, has category_name $cn;
$p isa product, has product_name $pn;
{ ($c, $p) isa category-assignment; }
or { not { ($c, $p) isa category-assignment; }; };
get $cn, $pn;
```

```
★ Fill in the Blanks ★
```

```
SELECT artist_name, album_name
FROM artists
FULL OUTER JOIN albums
ON artists.artist_id = albums.artist_id;
```

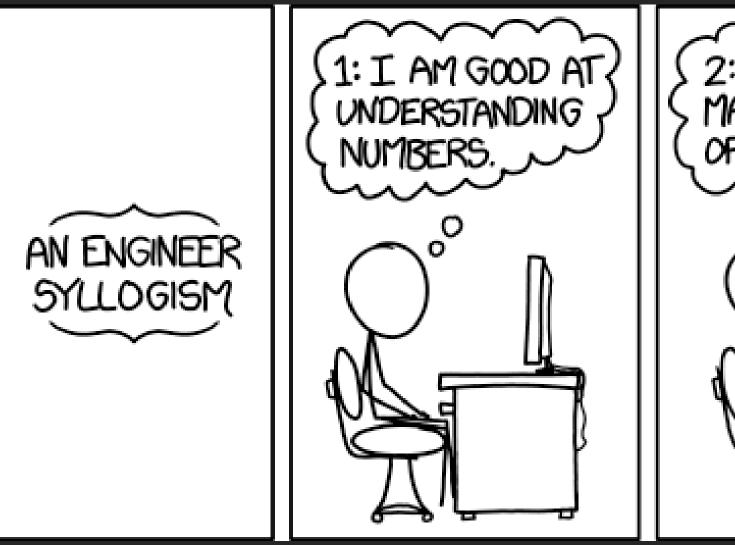
```
match
$ar isa artist, has artist_name $ar_name;
$al ___ album, has album_name $al_name;
{
    ($ar, $al) isa album-release;
} __ {
        ($ar, $al) isa ____;
};
get $ar_name, ____;
```

# Reading Data – Inference

- All men are mortal
- Socrates is a man
- Therefore, Socrates is mortal

```
men-are-mortal sub rule,
when {
    $man isa man;
},
then {
    $man isa mortal;
};
insert $m isa man, has name "Socrates";
match $mortal isa mortal; get;
```

# Reading Data – Inference







- All women are mortal
- Socrates is a man
- Therefore, Socrates is mortal

https://xkcd.com/1570/

# Reading Data - Inference

- When
  - Texas (\$b) is located in USA (\$a)
  - Dallas (\$c) is located in Texas (\$b)
- Then
  - Dallas (\$c) is located in USA (\$a)

```
transitive-location sub rule,
when {
   (location: $a, located: $b) isa locating;
   (location: $b, located: $c) isa locating;
}, then {
   (location: $a, located: $c) isa locating;
};
```

product_id	product_name	
14	Pepsi	$\overline{}$
15	Coca-Cola	
16	Fanta	
product_id	product_name	
		<b>↓</b> ↓
15	Coca-Cola	
16	Fanta	

```
product_name: Pepsi
has
product: V16576
$p
```

DELETE FROM products
WHERE product\_name = 'Pepsi';

match \$p isa product, has product\_name "Pepsi"; delete \$p;

```
★ Fill in the Blanks ★
```

```
FROM albums
USING artists
WHERE albums.artist_id = artists.artist_id
AND artists.artist_name = 'Michael Jackson'
AND albums.album_name = 'Bad';
```

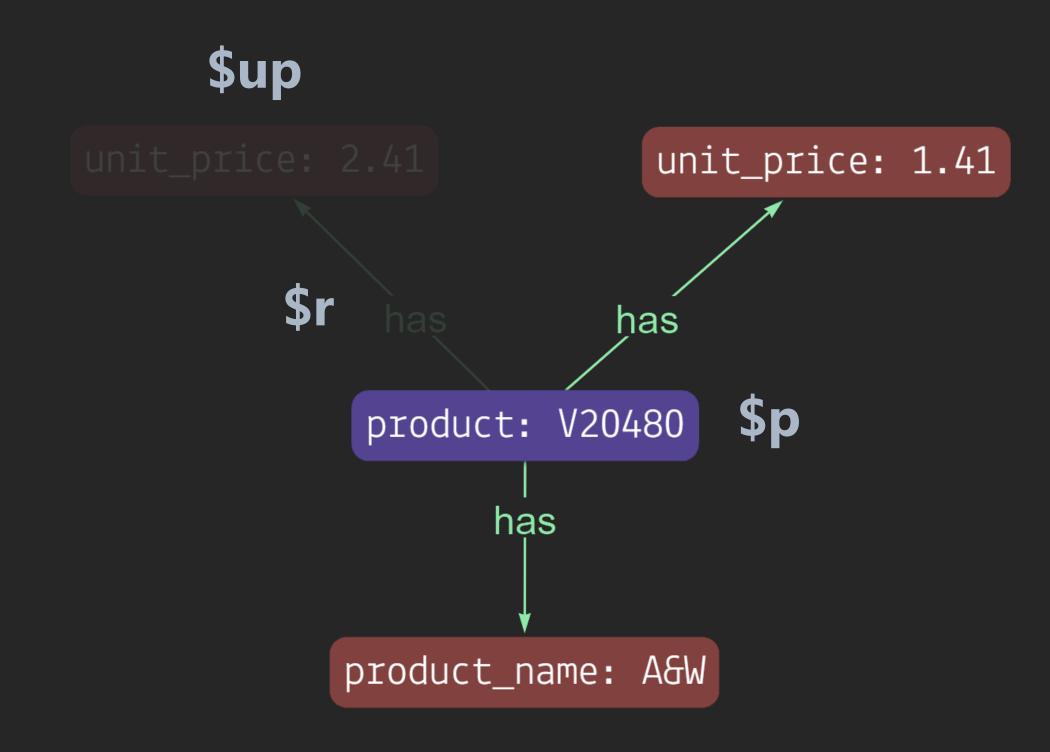
```
$\frac{1}{3}\arrisa \text{artist_name "Michael Jackson"; $\frac{1}{3}\arrisa \text{album_name "Bad"; $\frac{1}{3}\arrisa \text{album-release; }\frac{1}{3}\text{album-release; }\frac{1}{3}\text{album
```

product_id	product_name	unit_price
17	Sprite	\$1.55
18	7 UP	\$1.68
19	A&W	\$2.41

product_id	product_name	unit_price
17	Sprite	\$1.55
18	7 UP	\$1.68
19	A&W	\$1.41

```
UPDATE products
SET unit_price = 1.41
WHERE product_name = 'A&W';
```

### Graql



```
match $p isa product,
has product_name "A&W", has unit_price $up via $r;
delete $r;

match $p isa product, has product_name "A&W";
insert $p has unit_price 1.41;
```

UPDATE artists
SET artist\_name = 'Michael Joseph Jackson'
WHERE artist\_name = 'Michael Jackson';

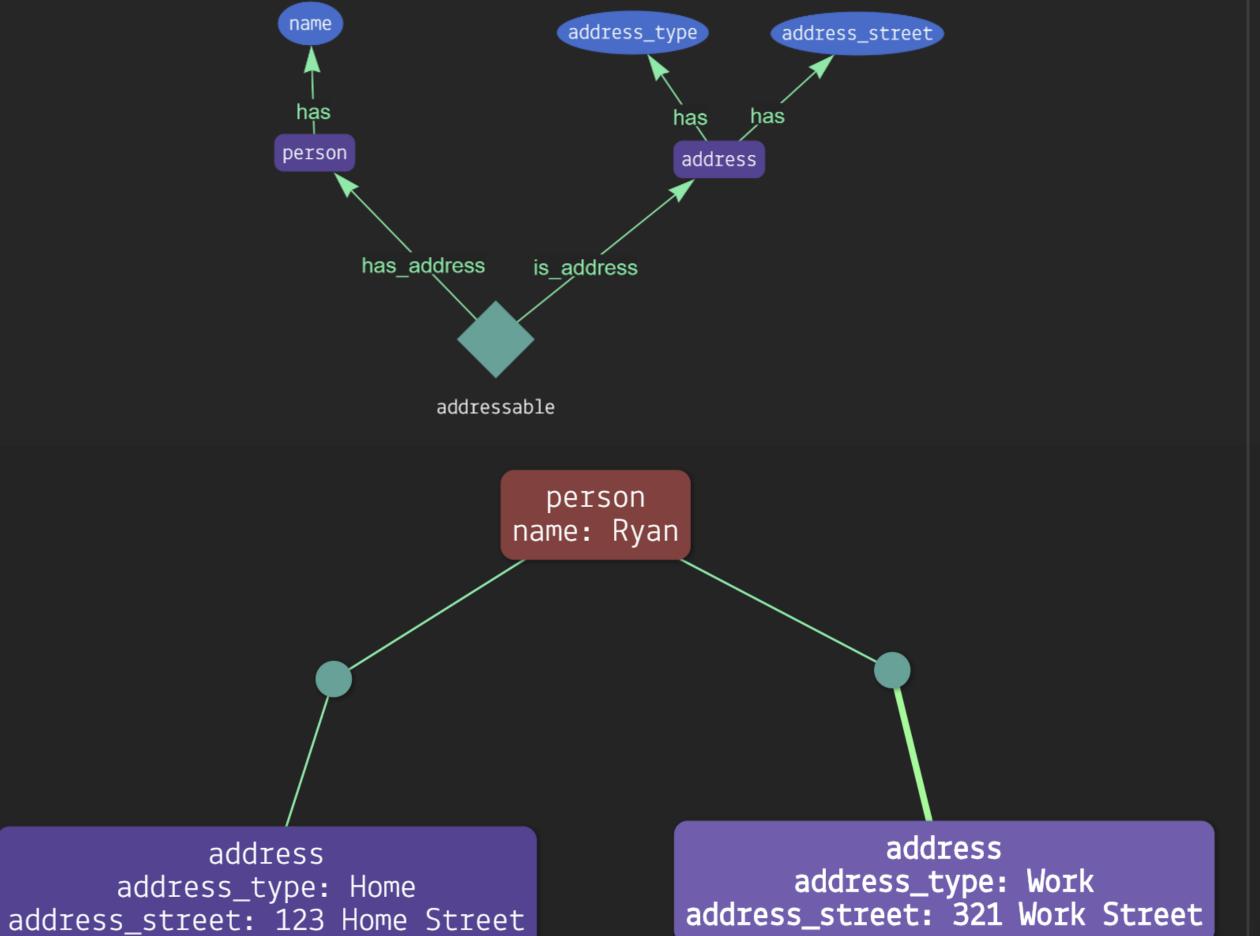
\_\_\_\_\_\$ar has artist\_name "Michael Joseph Jackson";
\_\_\_\_\_\$ar isa artist, has artist\_name "Michael Jackson" via \$r;
\_\_\_\_\_\$r;

\$ar isa artist, has artist\_name "Michael Jackson";

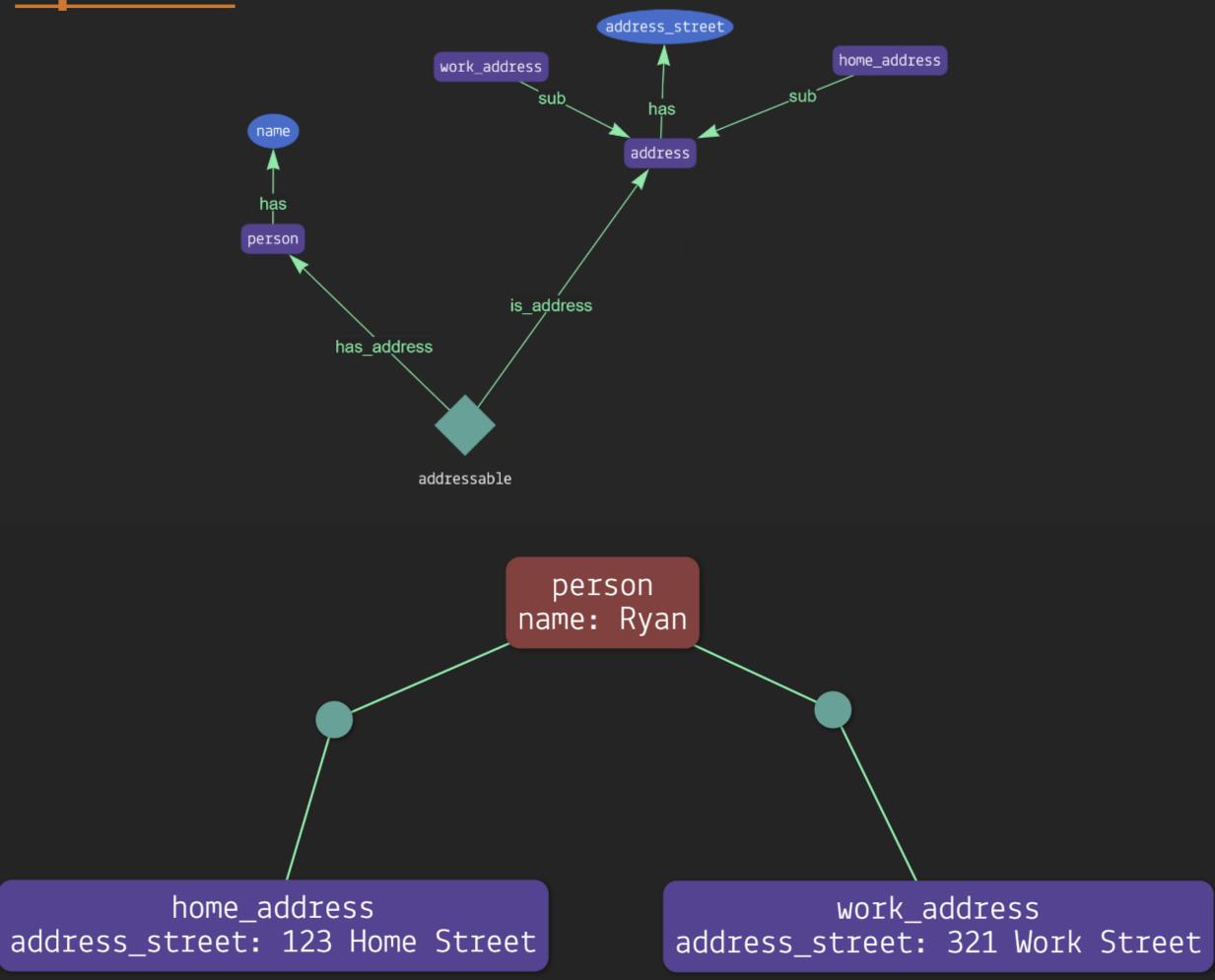
# Exercises

# Exercise – Which is Better?

### Option A



### **Option B**



# Exercise – Will This Work?

```
define
product sub entity,
 key product_name;
city sub entity,
 has city_name,
 plays location,
 plays located;
locating sub relation,
 relates location,
 relates located;
product_name sub attribute, datatype string;
city name sub attribute, datatype string;
insert
$bacon isa product, has product name "Bacon";
$toronto isa city, has city_name "Toronto";
(location: $bacon, located: $toronto) isa locating;
```

# Exercise - Convert Tables

#### **Objectives**

- 1. Convert the following tables to an equivalent Grakn schema
- 2. Populate the knowledge base with the equivalent data
- 3. Write a query to get the primary language of Argentina

		1	
DU	ni		les

country_id	country_name	population
1	Argentina	45,000,000

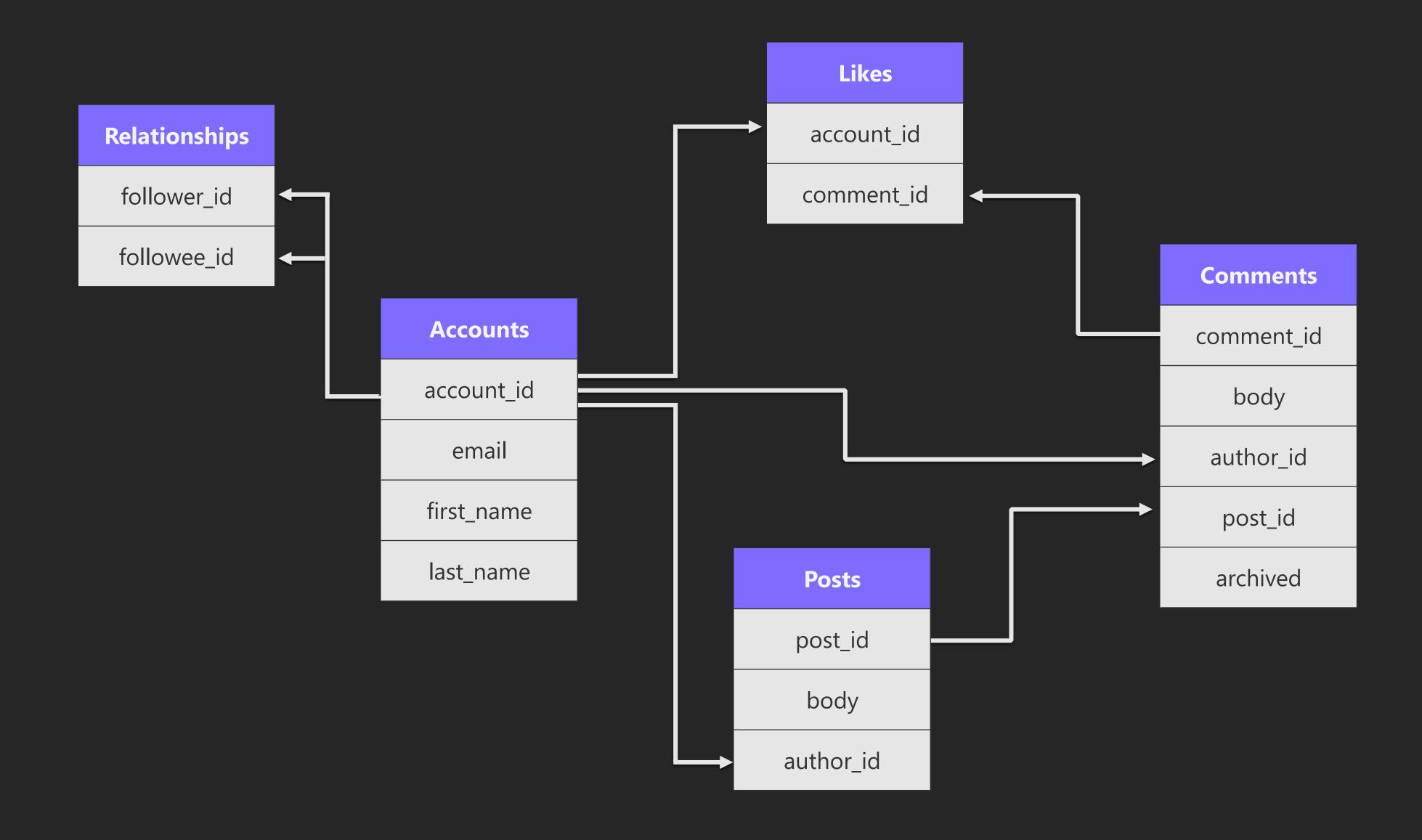
### CountryLanguages

language_id	country_id	primary
1	1	True
2	1	False

#### Languages

language_id	language_name	word_count
1	Spanish	150,000
2	English	600,000

# Exercise - Convert Schema



# Exercise – Linked List

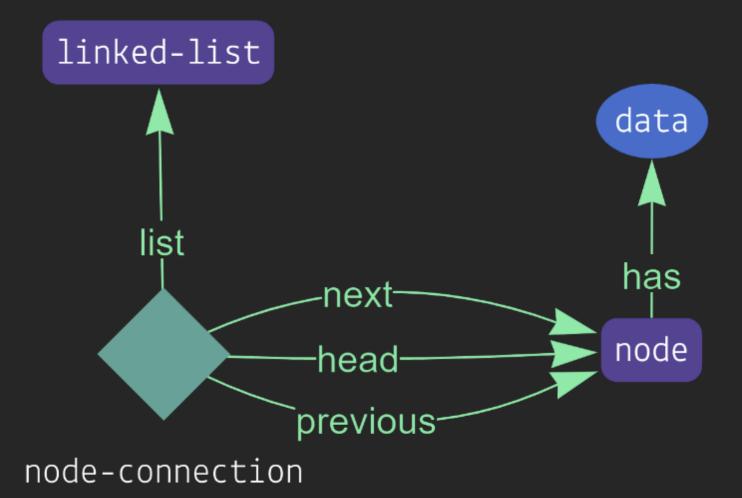
#### **Objectives**

- 1. Implement the given schema
- 2. Populate the knowledge base with the given data
- 3. Write queries to answer following questions:
  - 1. Who is in queue before Alexis?
  - 2. Who is in queue after the person after Bob?
  - 3. Who is the last person in the queue?

#### Data

Bob -> Sam -> Jessica -> Eric -> Alexis

#### Schema



# Exercise – Find Friends-of-Friends

### Objectives

- 1. Implement the given schema
- 2. Populate the knowledge base with the given data
- 3. Write a query to find the friends of Brandon's friends
- 4. Create a rule to infer two people may be acquaintances based on mutual friends

#### Data

Brandon's Friends:

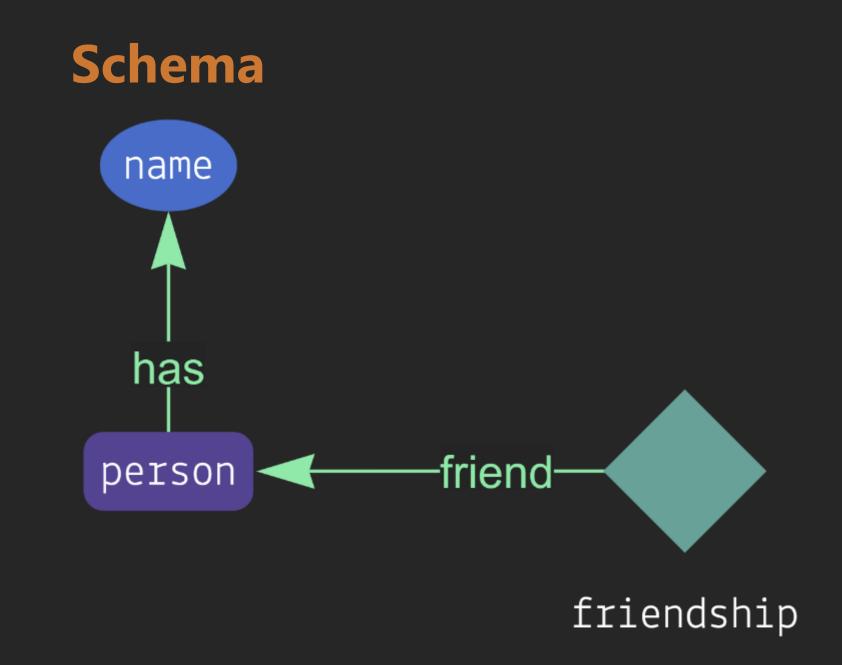
- Travis
- Mark
- Bob
- Melissa

Mark's Friends:

- Bob

Melissa's Friends:

- Louise
- Alice



# Exercise — Find Skilled Colleagues

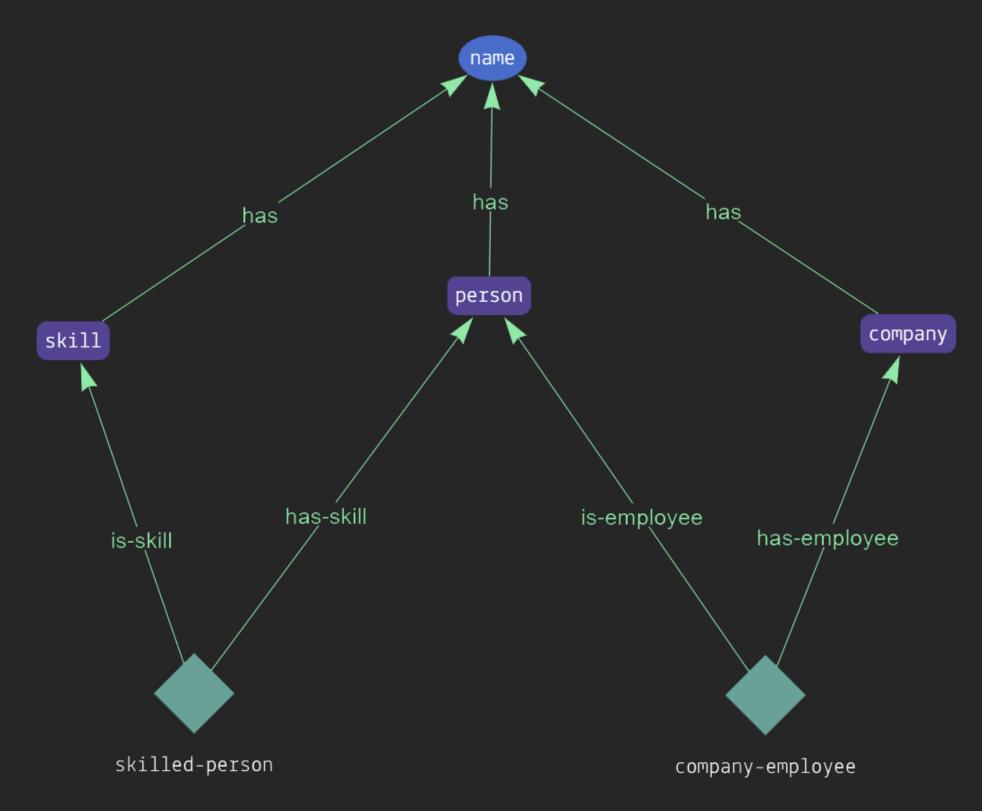
#### **Objectives**

- 1. Implement the given schema
- 2. Populate the knowledge base with the given data
- 3. Write a query to find everyone who is good at Kotlin and Java
- 4. Write a query to find everyone who has no experience with C#
- 5. Extend the graph to add the ability to assign skill level (weight)
  - 1. Update Eric to have a skill level of 3 in current skills
- 2. Update Larry to have a skill level of 2 in current skills
- 3. Update Sergey to have a skill level of 4 in current skills
- 4. Write a query to find everyone with a skill level greater than 2 in C#

#### Data

```
Company: Google
Employees:
Eric
Skills: Python, C#
Larry
Skills: C#, Kotlin, Java
Sergey
Skills: Kotlin, Java
```

#### Schema



# Questions/Comments?

### Resources:

- https://grakn.ai
- https://neo4j.com

