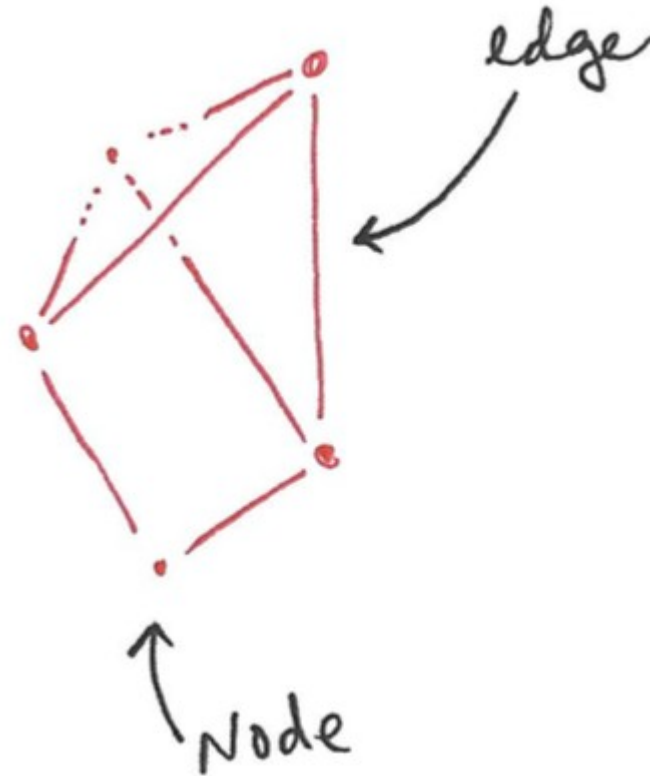


# Diseño y modelado en bases de datos de grafos



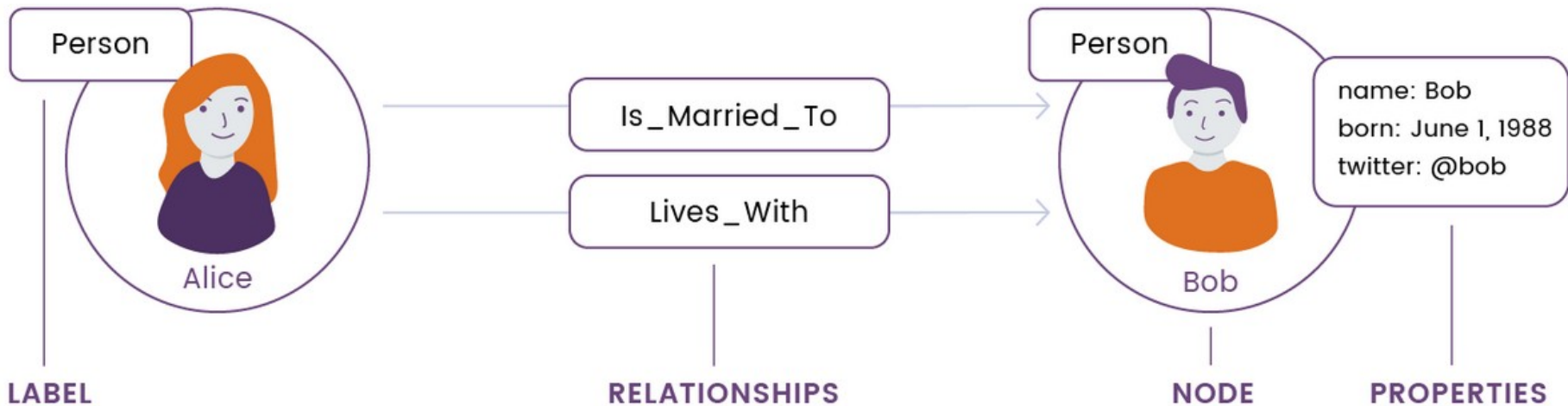
# Recordemos algunos conceptos

No hay un único modelo lógico para base de grafos

Los modelos más populares actualmente son ***property graph model (PGM)*** y **RDF + OWL**

**Nos concentraremos en PGM**

# Elementos del PGM



Las etiquetas representan los tipos o roles de los nodos

Las propiedades son parejas (clave, valor) y pueden estar asociadas a nodos y a relaciones (aristas).

# **El proceso de diseño**

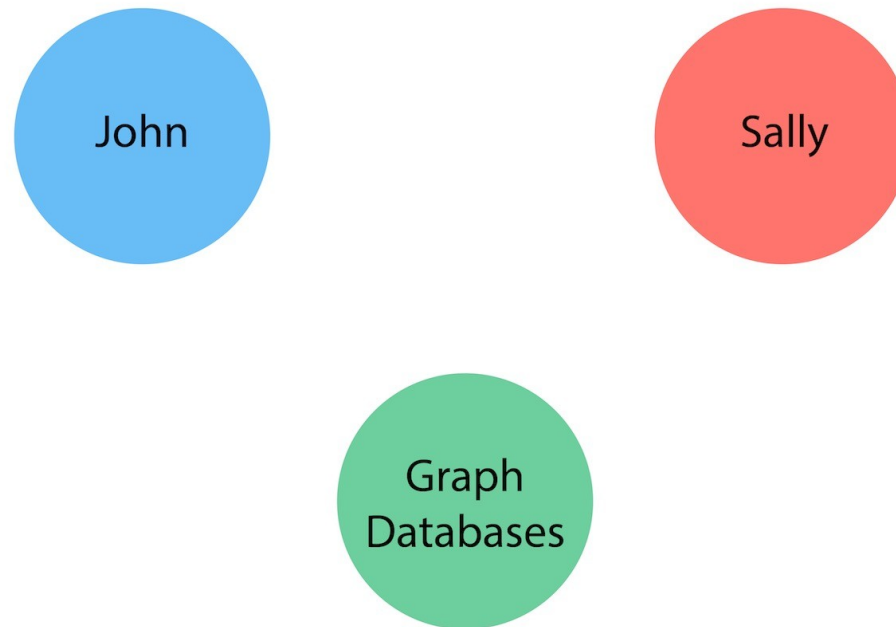
Consiste en decidir como representar a los elementos de la realidad en términos de nodos, etiquetas, relaciones y propiedades.

# Guías de diseño (1)

Nodos: usualmente representan **entidades** de la realidad.

Se pueden identificar nodos para el modelo de grafos a partir de sustantivos del dominio (ej: una persona, una empresa, etc.)

Two people, John and Sally, are friends. Both John and Sally have read the book, Graph Databases.



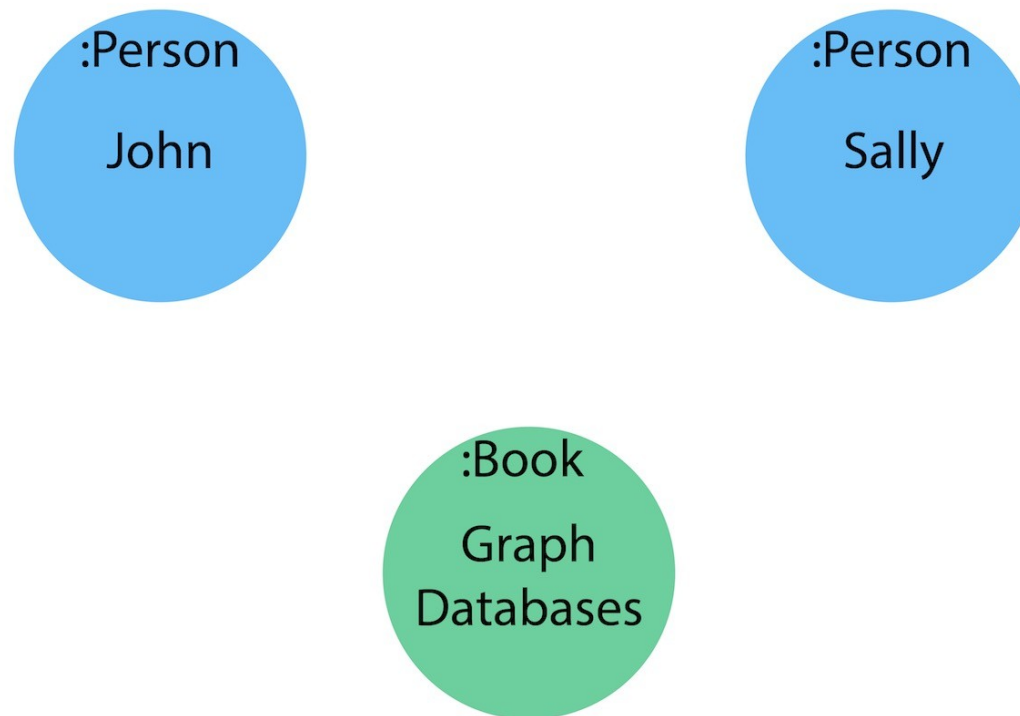
# Guías de diseño (2)

Etiquetas: se usan para agrupar nodos en conjuntos.

Dan una noción “liviana” de tipos.

Son opcionales, y cada nodo puede tener más de una etiqueta.

Two people, John and Sally, are friends. Both John and Sally have read the book, Graph Databases.



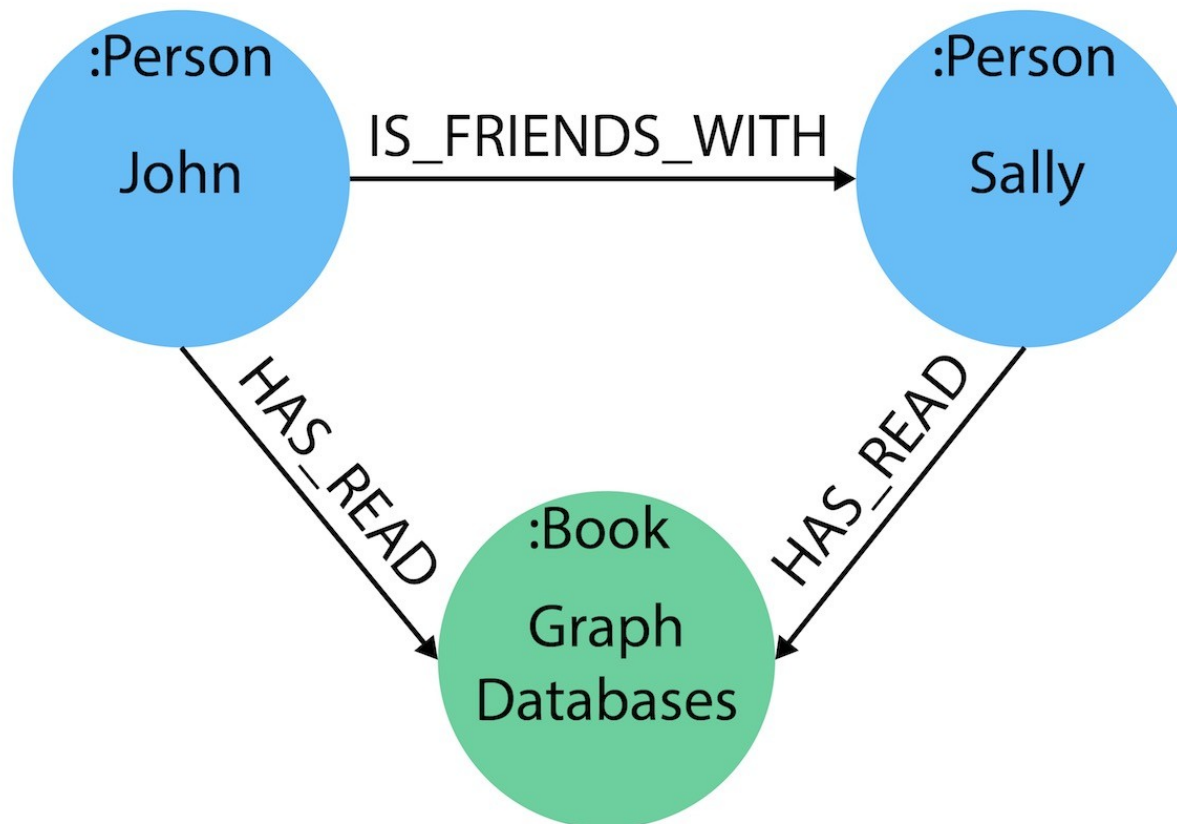


# Guías de diseño (3)

Relaciones: vinculan entidades (nodos) y tienen una dirección.

Se pueden identificar relaciones para el modelo de grafos a partir de verbos usados en el dominio (ej: dirige, conoce, lee, etc.)

Two people, John and Sally, are friends. Both John and Sally have read the book, Graph Databases.



# Guías de diseño (4)

Propiedades: permiten almacenar datos sobre las relaciones y los nodos.

Para identificarlos se puede partir de los atributos relevantes identificados en el dominio, y en las preguntas que se quieren responder

# Algunas preguntas posibles

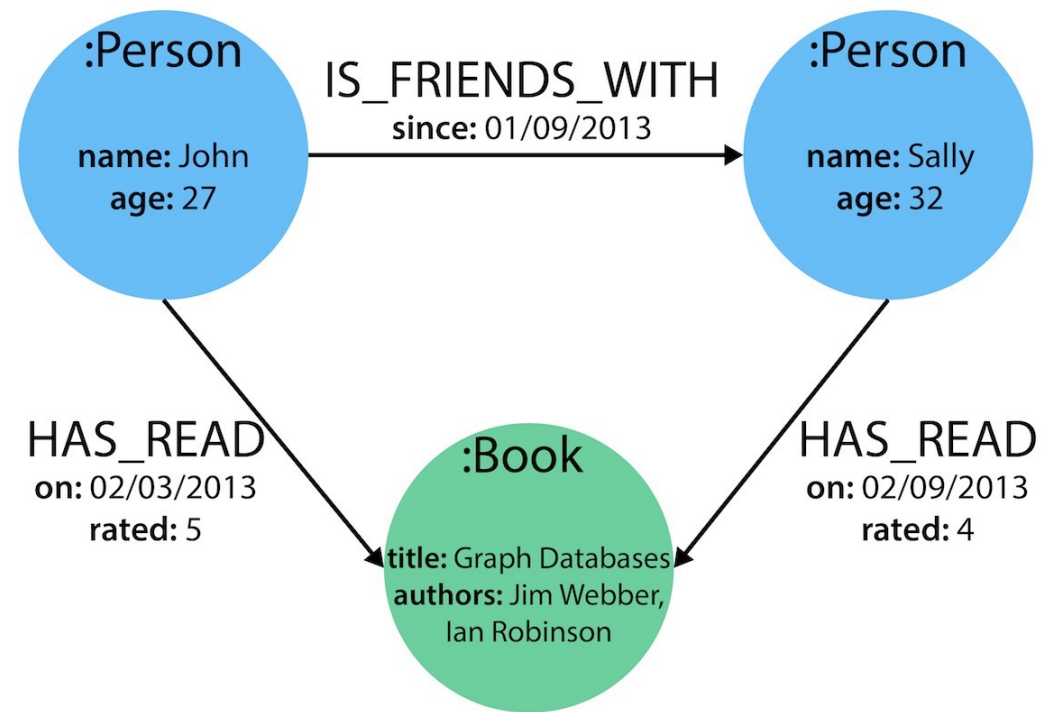
Cuánto hace que Sally y John son amigos?

Cuál es la calificación promedio del libro Graph Databases?

Quién es el autor del libro Graph Databases?

Cuál es la edad de Sally? Y la de John?

Quién leyó el libro Graph Databases antes, Sally o John?



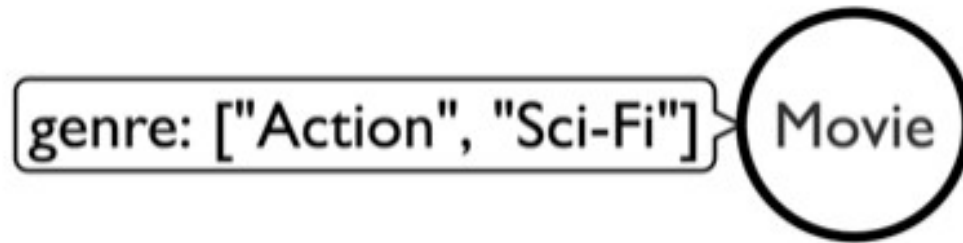
# Decisiones de diseño

Más allá de las recomendaciones generales algunos conceptos de la realidad tienen más de una representación posible.

¿Cuál elegir? Depende ...

Veremos a continuación algunas variantes y discusiones

# Propiedades vs. Relaciones



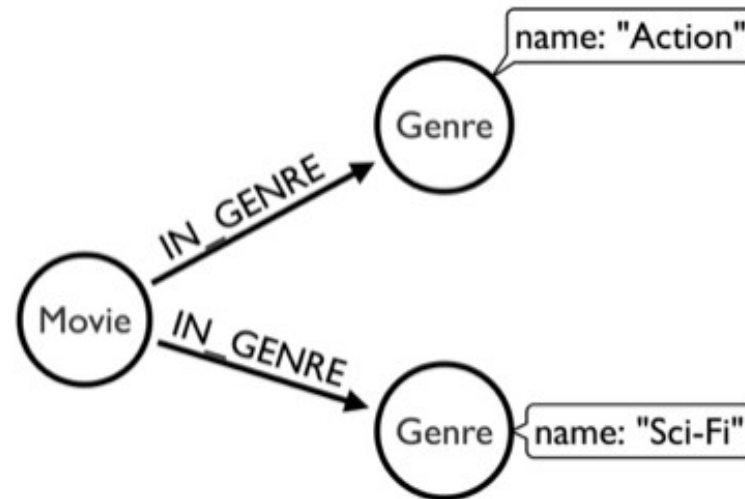
Devolver todos los géneros de una película

```
MATCH (m:Movie {title:"The Matrix"})  
RETURN m.genre;
```

Devolver todas las películas que comparten géneros

```
MATCH (m1:Movie), (m2:Movie)  
WHERE any(x IN m1.genre WHERE x IN m2.genre)  
AND m1 <> m2  
RETURN m1, m2;
```

# Propiedades vs. Relaciones (2)



Devolver todos los géneros de una película

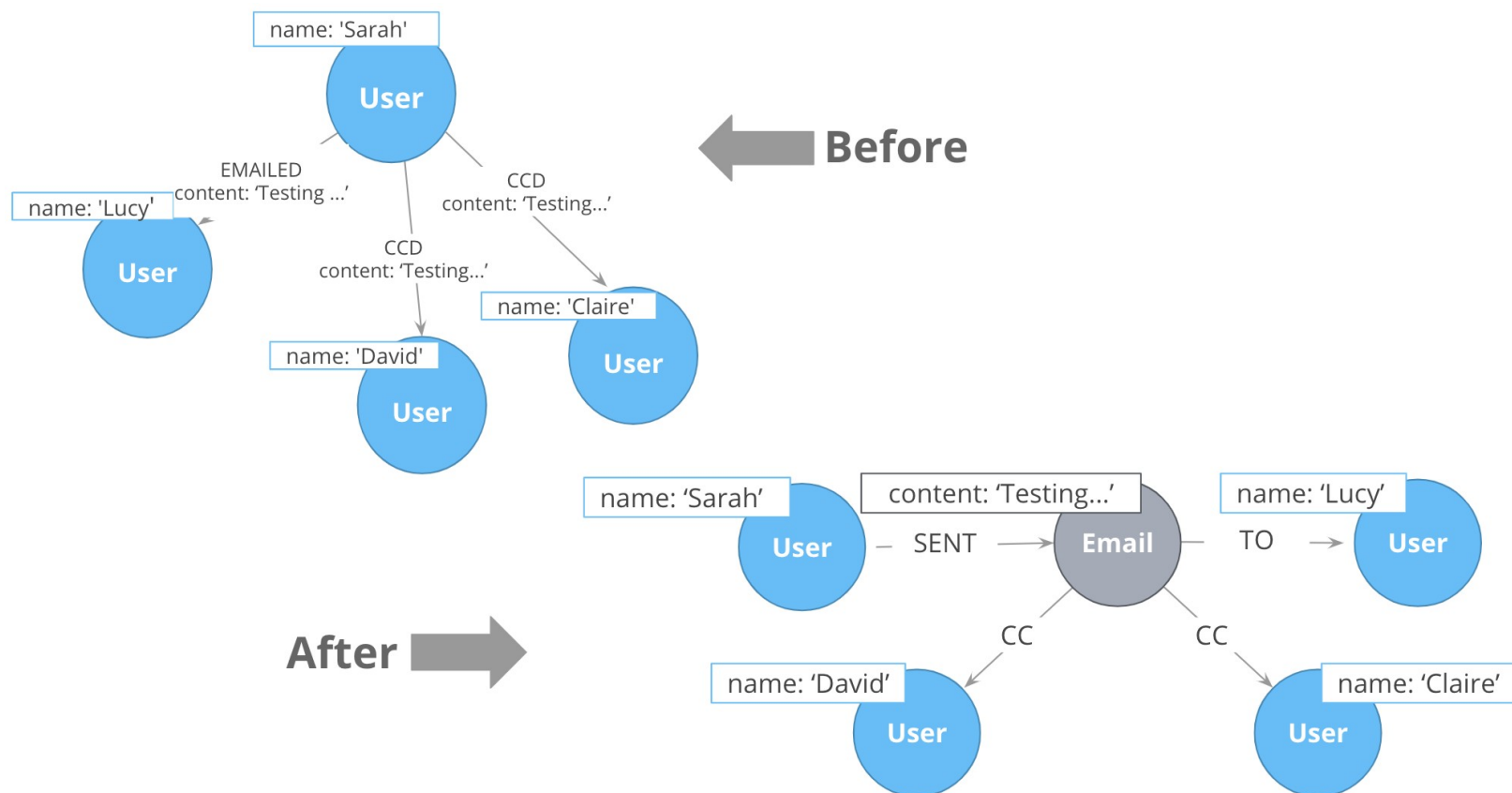
```
MATCH (m:Movie {title:"The Matrix"}),  
      (m)-[:IN_GENRE]->(g:Genre)  
RETURN g.name;
```

Devolver todas las películas que comparten géneros

```
MATCH (m1:Movie)-[:IN_GENRE]->(g:Genre),  
      (m2:Movie)-[:IN_GENRE]->(g)  
RETURN m1, m2, g
```

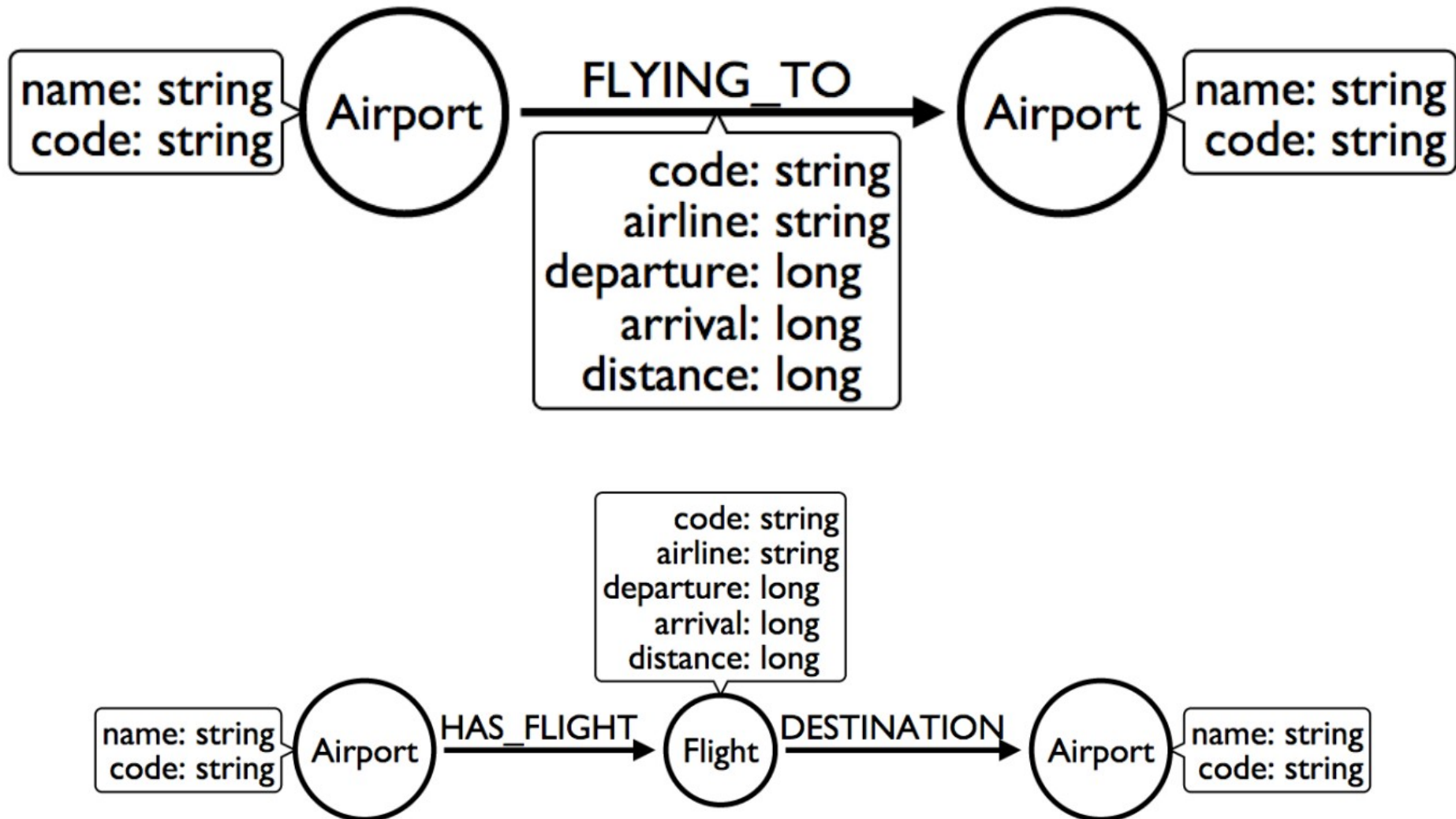
# Hiper aristas o nodos intermedios

Permiten representar relaciones entre más de dos entidades



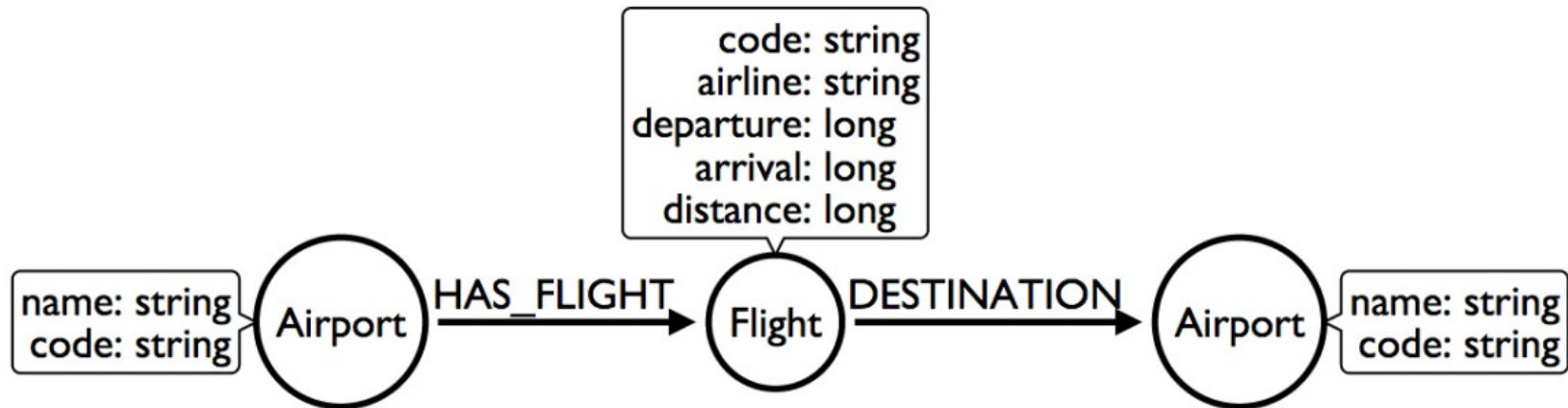


# Versionado y datos dependientes del tiempo



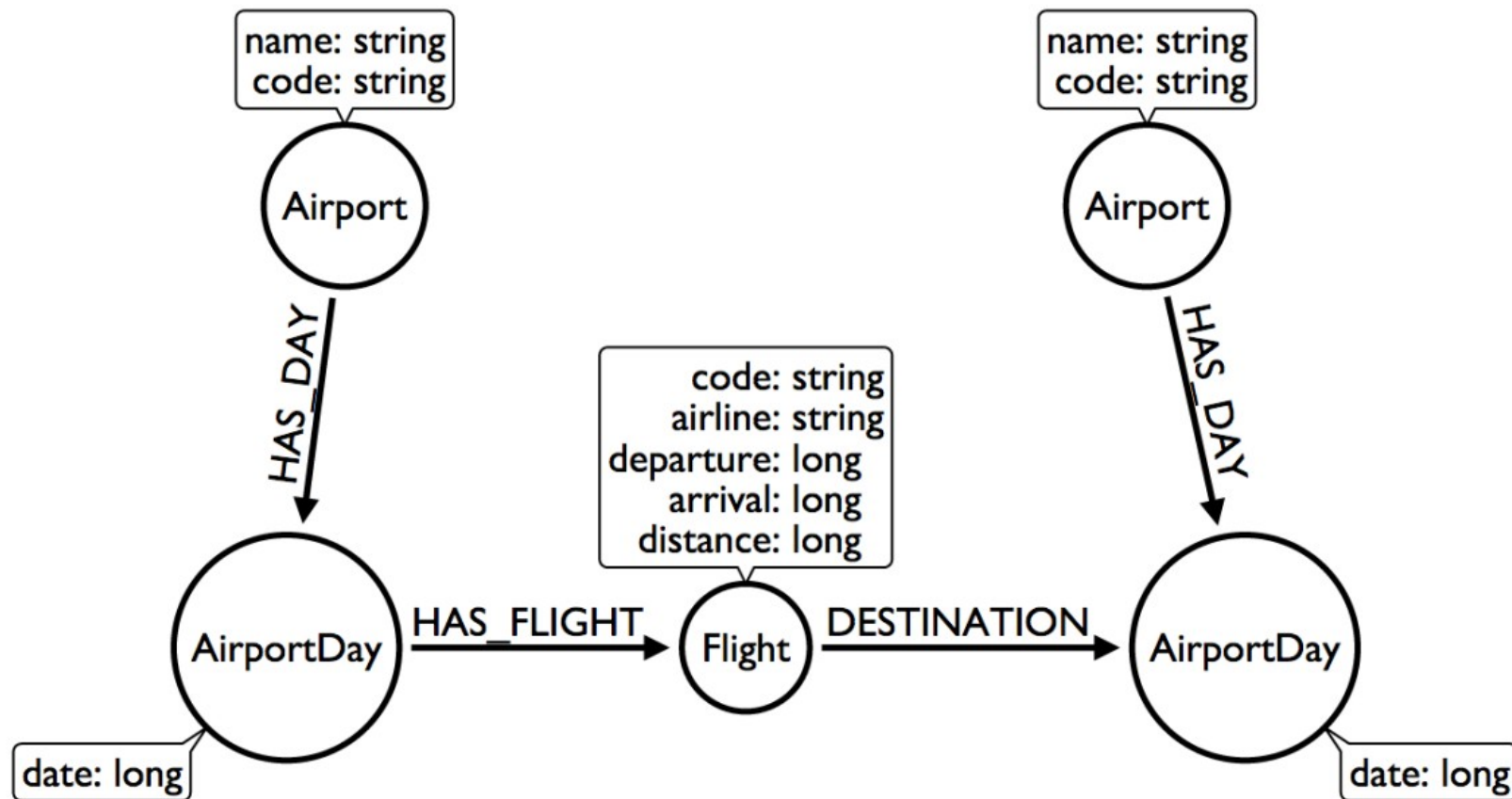
Supongamos que se quiere usar el grafo en una aplicación para búsqueda y reserva de vuelos

OBJETIVO: poder reducir rápidamente el subgrafo en que hago las búsquedas

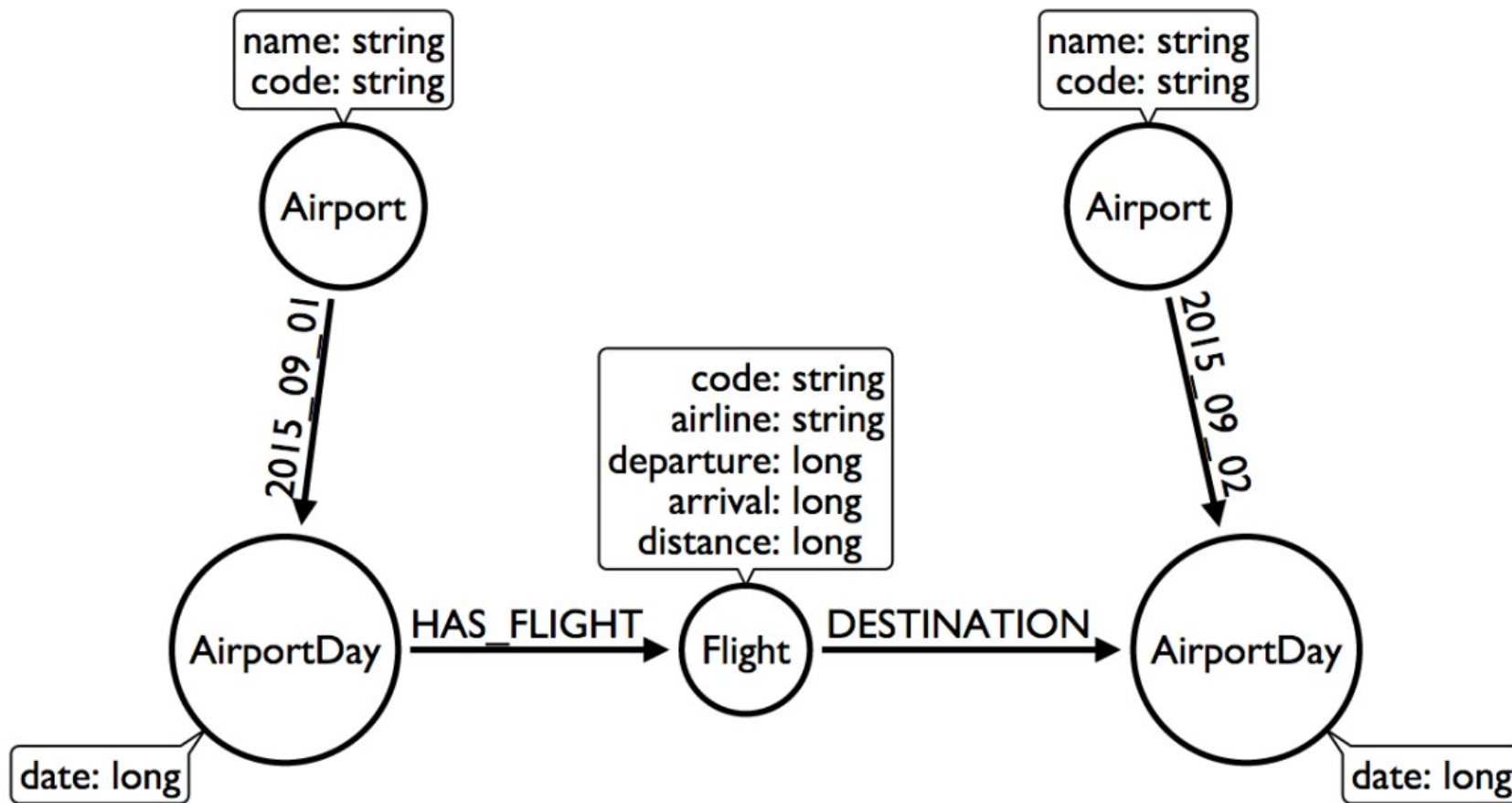


Sabemos el origen y el destino que interesa, y también la fecha del vuelo

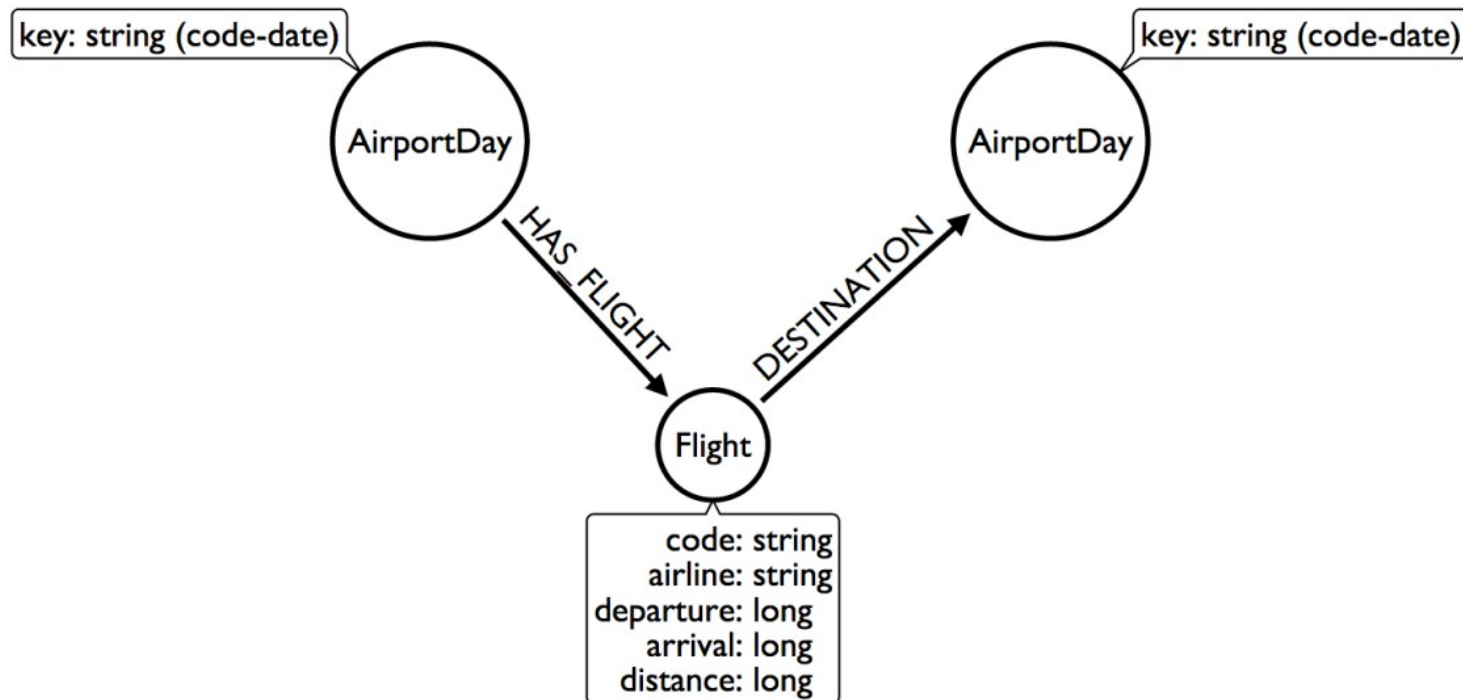
Sabemos el origen y el destino que interesa, y también la fecha del vuelo



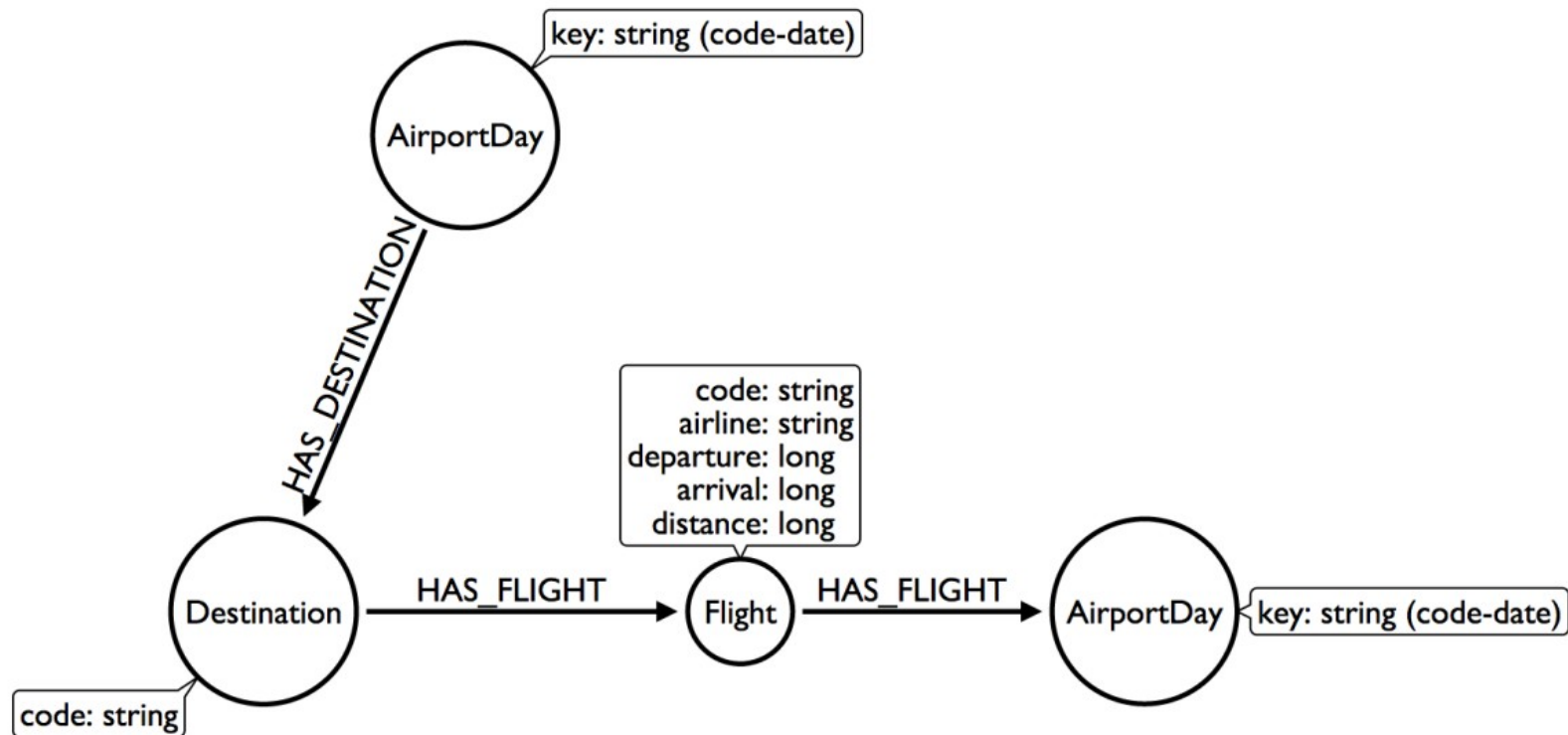
Sabemos el origen y el destino que interesa, y también la fecha del vuelo



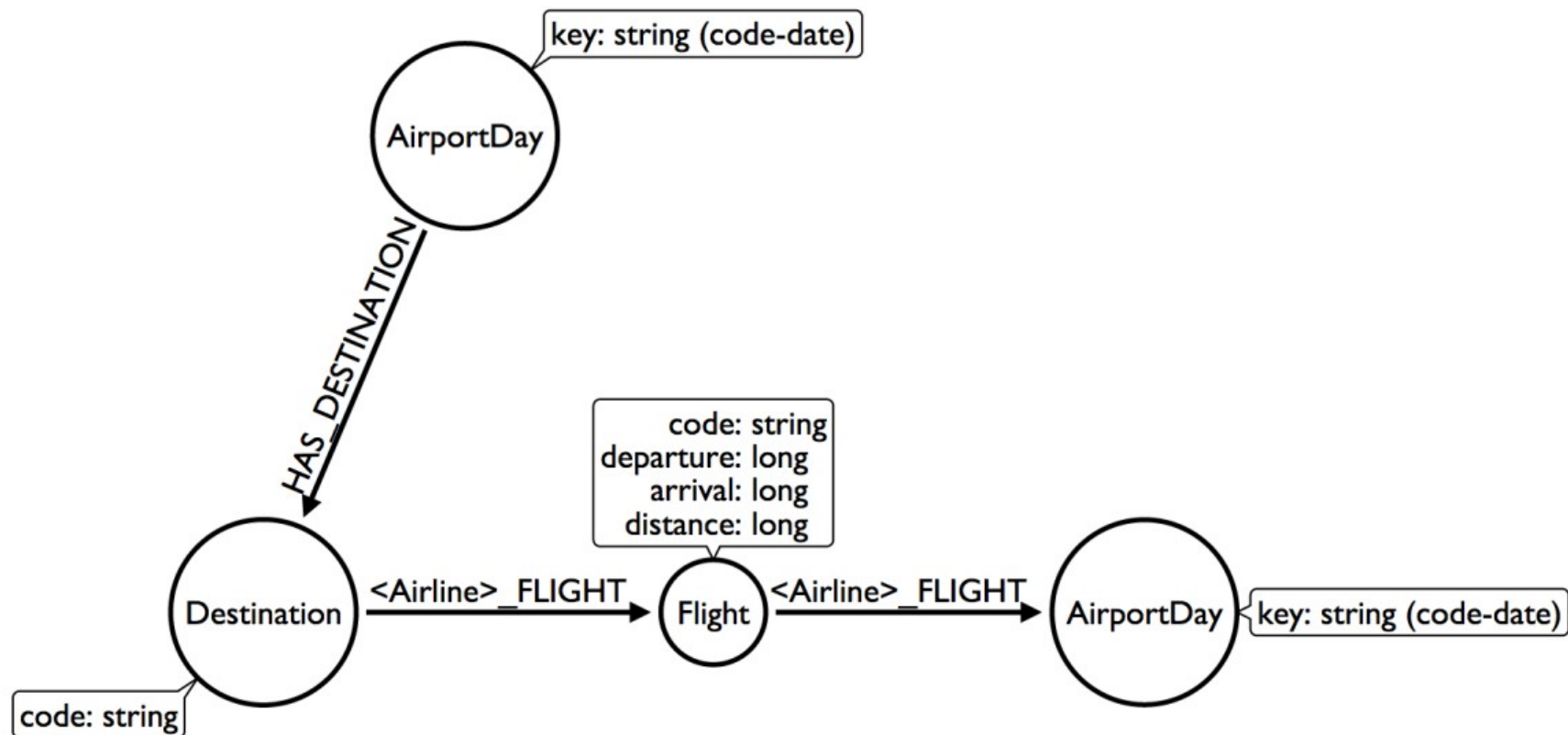
Podemos mejorar el acceso al AirportDay combinando el código del aeropuerto y la fecha (epoch)



Hay destinos que son muy populares (muchos vuelos).  
Podríamos mejorar aún más el modelo



Y si además queremos poder filtrar rápidamente por qué aerolínea opera el vuelo podríamos mejorar aún más el modelo



# ¿qué pasa si ya tengo un diseño relacional?

Se puede aplicar la estrategia de mapeo directo:

- Entidades a nodos
- Relaciones binarias a aristas
- Nodos intermedios para relaciones de 3 o más