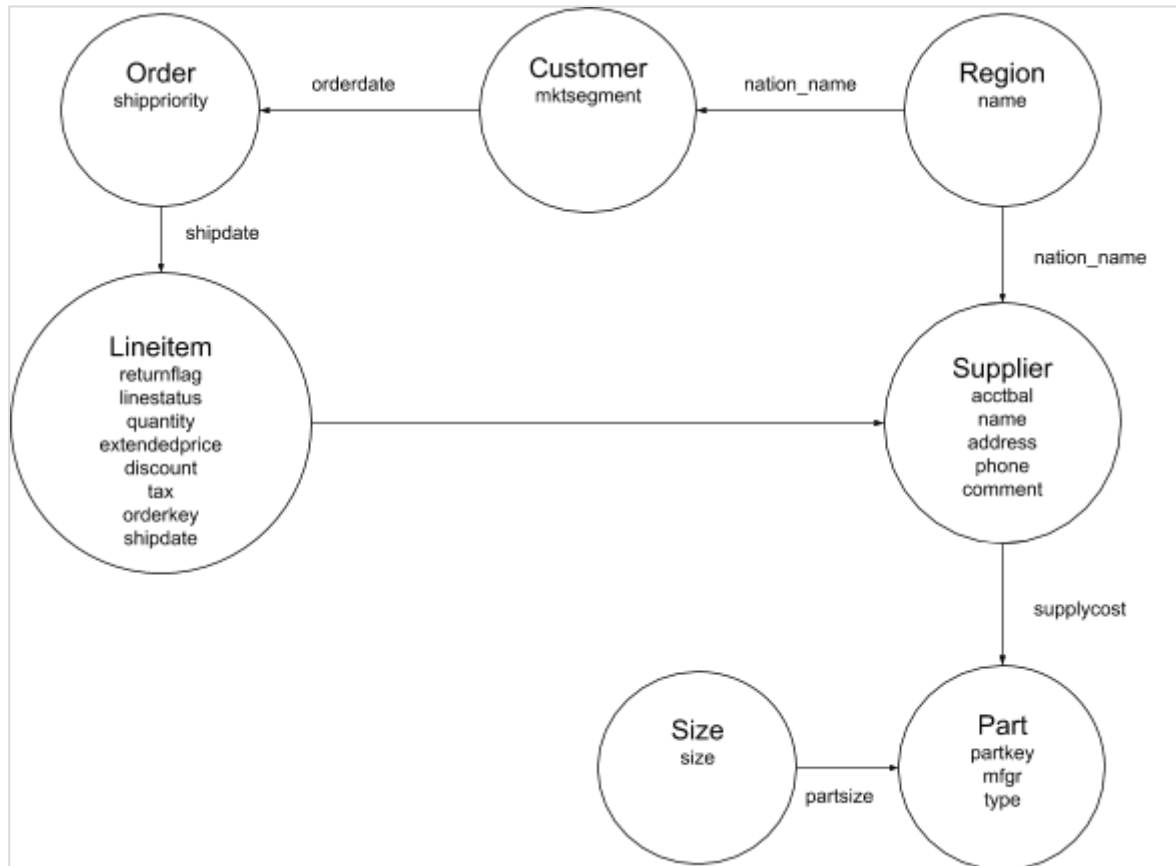


Graph databases with Neo4j

Omar Garcia Ríos
Jacobo Moral Buendía

Per dur a terme les quatre queries proposades sobre TPC-H, hem proposat el següent model de base de dades:



Model base de dades

A partir de les necessitats de cadascuna de les consultes, hem construït el model anterior, decidint que aquest és un model que s'ajusta a les necessitats de totes les consultes alhora.

Anàlisi de querys

Primera query:

La primera query es ben senzilla i només necessitem un node *Lineitem* que contingui totes les propietats o atributs que necessita la consulta. La nostra idea inicial era separar la propietat *shipdate* en un node separat. Tot i això, hem pensat la pèrdua d'espai no es compensaria amb la millora de temps, ja que la selecció és d'un rang en comptes d'un sol valor.

Segona query:

Aquesta segona consulta va ser més complicada alhora de buscar per on començar i a més de que a mesura que s'anava escrivint, ens vam trobar maneres alternatives d'implementar el model de bases de dades. Això ens va portar a prendre una decisió sobre les diverses alternatives que van anar sorgint. Però finalment ens vam decantar per una que creiem que es bastant correcte i dona uns resultats eficients.

El resum de la nostra idea es el següent: Començar per el node 'Size' per tal de descartar nodes al primer *match*, de manera que no fem lectures de nodes que no necessitem. Es per això que el tenim separat, per tal de filtrar informació que no ens interessa al primer pas de la query.

Una vegada tenim els nodes amb la 'size' correcta, ja es passa a fer un *match* que ens doni nodes amb l'estructura de relacions i tipus adequat. I seguidament un altre *match* que fem servir per obtenir el 'min_supply_cost' que s'ha d'utilitzar per fer la query.

Finalment si hem arribat a aquest punt, només s'ha de posar la clàusula *where* amb les condicions que marca la query original per tal de extreure la informació desitjada.

Tercera query:

Per a la tercera consulta, partim del node customer que compleixi la condició del select sobre el seu únic atribut *mktsegment*, des d'aquí naveguem per les seves relacions que també compleixin les restriccions de selecció.

Quarta query:

Per la quarta consulta, hem recorregut gairebé tots els nodes del model excepte *Part* i *Size*. Començant per la regió, hem anat recorrent les relacions per customer, al mateix cop que anem filtrant les condicions de la selecció, fins fer un cicle sencer.