**Explicació de l’Estructura dels Models de Dades**

En aquest document explicarem les decisions que hem pres per tal de fer que l’estructura dels models de dades sigui òptima segons les 4 queries donades. Hem dividit l’estructura en dues col·leccions: “Orders” (dissenyada per optimitzar les queries 1, 3 i 4) i “Partsupp” (dissenyada per optimitzar la query 2).

**ORDERS**

{

"o\_orderkey": 123,

"o\_orderdate": 2023-11-29,

"o\_shippriority": “Primary”,

"c\_mktsegment": “Cars”,

"nation": {

"n\_name": “Spain”,

"n\_regionkey": 1,

"r\_name": “Europe”,

"supplier": [

{

"s\_suppkey": 1093

}

]

},

"lineitems": [

{

"l\_suppkey": “Nissan”,

"l\_extendeprice": 1000,

"l\_discount": 50,

"l\_shipdate": 2023-11-14,

"l\_returnflag": ‘Y’,

"l\_linestatus": ‘S’,

"l\_quantity": 100,

"l\_tax": 133

}

]

}

(En aquest cas, sabem que posar *nation* com a embedded empitjorarà l’eficiència, però millora molt la comprensió del codi)

En primer lloc, vam començar a fixar-nos en la query 3, que accedeix a *orders*, *customer* i *lineitems*. Aquí, ens vam trobar amb el nostre primer dilema a l’hora de tractar la relació entre aquestes tres classes:

Si fem que *lineitems* englobi les altres dues, s’hauria de cercar la informació en molts documents diferents, ja que per buscar les dades d’una *order* hauríem de cercar en varis *lineitems* i, tenint en compte que *lineitems* té la multiplicitat més alta de totes, fa que aquesta no sigui l’opció ideal.

Si fem que *customer* englobi les altres dues, tampoc seria una bona idea. Al cap i a la fi, el que més es crea en aquesta database són *orders,* llavors estant englobades dins de *customer,* es podrien generar problemes de padding, ja que el document podria créixer sovint i MongoDB no deixaria l’espai reservat suficient.

Per tant, creiem que la millor idea és que *orders* englobi les altres dues classes, evitant aquests problemes i, a més, fent que l’agrupació de la query sigui òptima, ja que tots els paràmetres del GROUP BY estan dins d*’orders.* A més, hem posat un índex a *c\_mktsegment* per tal de millorar l’eficiència de la cerca a l’hora de fer el WHERE de la query.

En segon lloc, vam fixar-nos en la query 1, que només vol accedir a informació de *lineitems,* cosa que fa senzilla la seva integració amb l’esquema explicat anteriorment. A més, hem posat un índex a *l\_shipdate* per a millorar l’eficiència del WHERE de la query.

Finalment, en tercer lloc, vam fixar-nos en la query 4, per a nosaltres la més difícil d’integrar. En aquesta query es vol accedir a informació de *customer, orders, lineitems, supplier, nation* i *region;* per tant, havíem d’integrar les tres últimes amb l’esquema anterior.

Per a fer-ho, vam decidir agrupar *supplier* i *region* dins de *nation,* ja que aquesta era el nexe entre les classes i, així, es facilitava l’agrupament de la query (el GROUP BY és per *n\_name*).

A més, hem posat un array amb tots els *suppliers* de cada *nation*, per tal de tenir tots els *suppliers* agrupats per *nations*, cosa que fa que el join amb *customer*, *nation* i *supplier* sigui més ràpid, eficient i més fàcil. També hem posat indexos a *n\_regionkey* i a *r\_name* per tal de millorar la rapidesa del WHERE de la query.

**PARTSUPP**

{

"ps\_partkey": 20201102112,

"ps\_suppkey": 1093,

"ps\_supplycost": 15000,

"s\_acctbal": 100,

"s\_name": “Nissan”,

"s\_address": “17-1 Ginza 6-Chome Chuo-Ku Tokyo, Japan”,

"s\_phone": “0120-315-232”,

"s\_comment": “Somos un grupo empresarial de tradición y líder del mercado que aporta bienestar económico y social a todos nuestros clientes, empleados y proveedores.”,

"n\_name": “Japan”,

"r\_name": “Asia”,

"p\_mfgr": “MICHELIN PRIMACY 4 195/65 R15 91 V”,

"p\_size": 125,

"p\_type": “Tyres”,

}

En aquest cas, vam decidir fer una altra col·lecció dissenyada per optimitzar la query 2, ja que trobàvem que aquesta era molt disjunta a les de més i, per tant, no hi havia sentit intentar ajuntar-ho tot en una.

En la query 2 s’accedeix a les classes *part, supplier, partsupp, nation* i *region.* Vam decidir englobar les altres classes a *partsupp,* ja que des de *partsupp* pots accedir a les altres classes (a *part* i a *supplier* mitjançant *ps\_partKey* i *ps\_suppkey*, a *nation* mitjançant *s\_nationkey* i a *region* mitjançant *n\_regionkey*). Així, tot el document pot ser un document pla amb els atributs necessaris, millorant l’eficiència. Finalment, hem posat un índex en *r\_name* per tal que la condició del WHERE de la query sigui més ràpida.