

Núria Bruch

Florencia Rímolo

Query 1

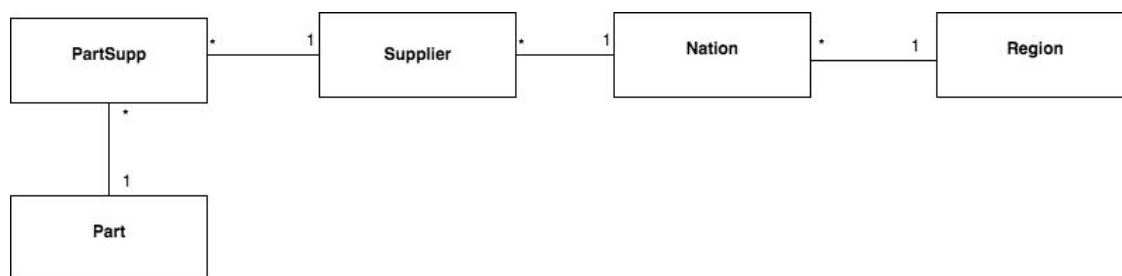
```
{ "_id": "l_returnflag" + " " + "l_linestatus"  
  "lineitems": [  
    "l_quantity": "",  
    "l_extendedprice": "",  
    "l_discount": "",  
    "l_tax": "",  
    "l_shipdate": "", ... ] }
```

En aquesta query, veiem que només fa servir una taula: Lineltem. Per tant, el que hem fet és guardar com a id els atributs del GROUP BY, i després els que necessitem per calcular els atributs d'agregació. Ja que, si ens guardem els atributs ja agrupats, en un UPDATE o INSERT podria passar que hauríem de recalculer-ho tot.

```
{ "_id": l_orderkey + " " + l_linenumbr,  
  "l_returnflag": "",  
  "l_linestatus": "",  
  "l_quantity": "",  
  "l_extendedprice": "",  
  "l_discount": "",  
  "l_tax": "" }
```

Aquesta és la segona versió que hem proposat com a document de la primera query: hem canviat el id i hem reorganitzat els atributs amb l'objectiu de poder ajuntar el document amb la resta de documents de les altres queries més fàcilment.

Query 2



```
{ "_id": "ps_partkey" + " " + "ps_suppkey",  
  "part": { "p_mfg": "",  
    "p_size": "",  
    "p_type": "" },  
  "supplier": { "s_acctbal": "",  
    "s_name": "",  
    "s_address": "",  
    "s_phone": "" },
```

```

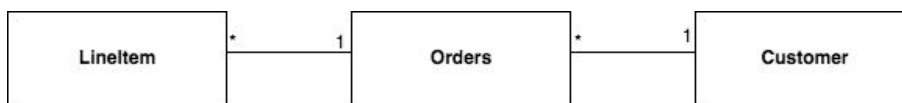
"s_comment": "",
"n_name": "",
"r_name": ""},
"ps_supplycost": ""}

```

En aquesta query hem tingut dificultats en trobar la id ja que no hi ha GROUP BY. Hem fet l'esquema relacional per ajudar-nos, i ens hem fixat que, si agafem la id del PartSupplier, navegant a qualsevol direcció la cardinalitat és 1 i, per tant, no necessitem fer cap llista.

Així doncs, hem guardat els atributs que es necessiten al SELECT i els atributs de les condicions on ens passem per paràmetre el valor a comparar. Hem decidit agrupar els atributs de part i supplier en dos subdocuments per tal de tenir els elements més organitzats, tot i que no és necessari.

Query 3



```

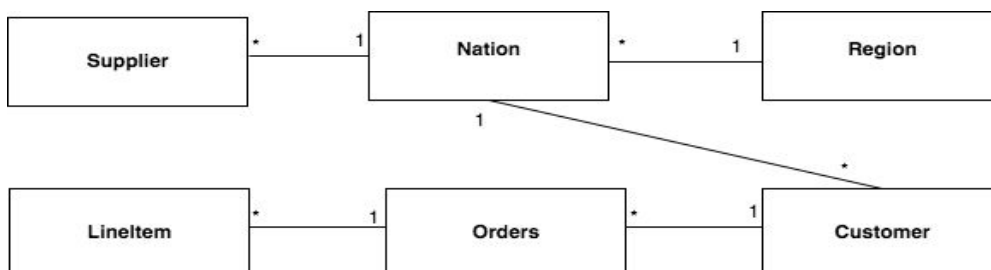
{"_id": "l_orderkey",
"o_orderdate": "",
"o_shippriority": "",
"c_mktsegment": "",
"lineitems": [
  {"l_shipdate": "",
   "l_extendedprice": "",
   "l_discount": ""}
]}

```

Per aquesta query, hem decidit la id mirant els atributs del GROUP BY. Primer havíem pensat que la id fos orderkey + orderdate + shippriority, però com que els dos últims depenen del primer que es la PK de la taula, hem decidit separar-los de manera que la id sigui només el l_orderkey i després afegir els altres dos com a atributs.

Després teníem dos opcions: afegir un atribut revenue que seria el resultat de calcular $\text{sum}(l_extendedprice * (1 - l_discount))$, o guardar els atributs l_extendedprice i l_discount per calcular el revenue a partir d'aquests. Seguint el mateix argument que a la query 1, hem optat per la segona opció.

Query 4



```

{"_id": "l_orderkey" + " " + "l_linenum",
"l_extendeprice": "",

```

```

    "l_discount": "",
    "o_orderdate": "",
    "r_name": "",
    "n_name": "",
    "c_nationkey": "",
    "s_nationkey": ""}

```

Per aquesta última query, primer vam pensar en posar com a id l'atribut n_name. Però després ens vam adonar que fent-ho així l'estructura del document seria massa complexa ja que estaríem agrupant molts objectes que tinguin aquell n_name, ja que no és la PK de la taula i que la cardinalitat agafant aquest atribut com a id seria massa gran.

Ajudant-nos de l'esquema relacional, hem decidit que la id podria ser la de la taula Lineitems, ja que totes les navegacions a partir d'aquesta taula tenen cardinalitat 1. Això mirant l'esquema podria no semblar cert ja que la cardinalitat de Suppliers és * però, si ens fixem en la query es fa una join de l_supplekey = s_supplekey.

També remarcar que els dos últims atributs que hem posat són necessaris perquè fa una join a partir d'aquests atributs que, entre ells, no són FK.

Collections

lineitem:

```

{
  "_id": l_orderkey+"_"+l_linenumbr,
  "l_orderkey": "",
  "l_returnflag": "",
  "l_linestatus": "",
  "l_quantity": "",
  "l_extendedprice": "",
  "l_discount": "",
  "l_tax": "",
  "l_shipdate": "",
  "o_orderdate": "",
  "o_shippriority": "",
  "c_mktsegment": "",
  "c_nationkey": "",
  "s_nationkey": "",
  "r_name": "",
  "n_name": ""
}

```

partsup:

```

{
  "_id": ps_partkey+"_"+ps_supplekey,
  "p_partkey": "",
  "p_mfgr": "",
  "p_size": "",
  "p_type": "",
  "s_acctbal": "",
  "s_name": "",
  "s_address": "",
  "s_phone": "",
  "s_comment": "",
  "n_name": "",
  "r_name": "",
  "ps_supplycost": ""
}

```

El primer document serveix per poder resoldre les queries 1, 3 i 4 ja que les tres tracten principalment atributs de lineitem que es poden ajuntar en un mateix document.

Per acabar, com que la única query que no podem resoldre amb el document anterior és la 2, tindrem un document igual que el que havíem dissenyat prèviament per poder resoldre la query 2.

Un cop fent les queries, ens hem adonat que al document lineitem ens feia falta afegir la orderkey per tal de poder fer correctament la query 3, ja que al tenir aquest atribut però a la id del document a l'hora de fer la query no trobava l'atribut. El mateix passa en el document partsup i l'atribut p_partkey, que és necessari per resoldre correctament la query 2 i que en un primer plantejament no guardàvem.