

Conceptes per a Bases de Dades Especialitzades

Document Stores: MongoDB

En primer lloc, va ser necessari descarregar MongoDB i familiaritzar-nos amb el seu entorn i la shell. Vam seguir un tutorial bàsic per a descobrir les àmplies característiques amb les que compta. Després de fer això, vam decidir que utilitzaríem l'API de Python (PyMongo) per a dur a terme la nostra pràctica. A partir d'aquest moment vam començar a analitzar com modelar les taules en documents per a així poder decidir una estructura eficient de les col·leccions a utilitzar adequant-nos a les queries. Un com vam haver entès les relacions entre les diverses taules, ens vam enfocar en analitzar les quatre queries i els elements que combinen. A continuació, comentarem query per query les conclusions que vam anar extraient a cada pas:

QUERY 1

En aquesta primera query solament es treballa amb la taula Lineitem, amb la qual cosa vam pensar que seria una peça molt important dins de la modelització, però no ens donava molta informació sobre com organitzar les altres taules entre elles.

QUERY 2

En aquest cas, es combinen les taules Part, Supplier, Partsupp, Nation i Region. Ràpidament vam pensar que la informació de Region es podria afegir dins de l'objecte Nation, ja que Region solament té un atribut (name). Aquest join uneix les diferents taules mitjançant les foreign keys i ademés té en compte els atributs type i size de Part, l'atribut name de Region, i l'atribut supplycost de Partsupp. Per altra banda, la gran majoria dels atributs a seleccionar són de les taules Supplier i Part. Això ens va fer pensar que la taula principal per a aquesta query seria Supplier i que probablement podríem incloure un array de Partsupp en el seu interior, i que en el mateix moment, aquest inclogués algun atribut necessari de Part. Aquesta col·lecció també inclouria Nation, i com ja hem comentat abans, dins d'aquest, tindríem Region. També es veu clarament que no té cap taula en comú amb la query anterior.

QUERY 3

En aquesta query es tenen en compte les taules Customer, Orders i Lineitem. Tal i com utilitza les taules la query 3, vam pensar que la millor manera de representar-ho mitjançant document stores és utilitzar la taula Lineitem com a principal i que dins contingui la teula Orders, i a la vegada, a l'interior d'aquesta, estigui continguda la taula Customers. A més, es pot veure clarament que la majoria d'atributs que utilitza són de Lineitem, i algun d'Order. També vam recordar la Query 1 que tractava exclusivament Lineitem, i per tant, la decisió d'utilitzar Lineitem com a base de la col·lecció ens va semblar la més adequada.

QUERY 4

En aquesta última query s'utilitza Customer, Orders, Lineitem, Supplier, Nation i Region. A diferència de les queries anteriors on havíem vist una clara separació entre les taules Customer, Orders, Lineitem, i per altra banda, Supplier, Nation i Region, on aquestes tractaven també amb Partsupp i Part, en aquest cas, veiem una unió entre aquestes. Tot i això, vam decidir mantenir la idea d'utilitzar dues col·leccions, amb Lineitem i Supplier com a

bases, i que aquestes incloessin les respectives taules comentades. En aquest punt vam decidir incloure a la col·lecció Lineitems l'atribut suppnationkey, per representar la nation key del Supplier, i a més, la nation key, nation name i la region name del Customer. També en aquest punt ens vam plantejar utilitzar una sola col·lecció per evitar problemes en el join, però vam pensar que seria contraproductiu per a la resta de queries.

En conclusió, després d'analitzar detingudament cada una de les queries i l'estructura de les taules, hem decidit organitzar la document store utilitzant dues col·leccions, de la següent manera:

Col·lecció 1:

//in Lineitem

```
{
  "id": 1,
  ...
  "suppnationkey": 1,
  "order":
  {
    "orderdate": 15/03/2020,
    ...
    "customer":
    {
      "mktsegment": segment1
      "nationkey": 2,
      "nationname": "Francia",
      "regionname": "EU"
    }
  }
}
```

Col·lecció 2:

//in Supplier

```
{
  "id": 1,
  ...
  "nation":
  {
    "name": "Spain",
    ...
    "region":
    {
      "name": "EU",
      ...
    }
  }
  "partsupp": [
    {
      "supplycost": 50,
      ...
      "part":
      {
        "size": 3,
        ...
      }
    }
  ]
}
```

Per últim, hem creat els següents índexs ja que hem pensat que seran útils per a millorar l'eficiència de l'execució en cadascuna de les queries de l'1 al 4 respectivament:

```
db_lineitems.create_index("shipdate")
db_lineitems.create_index("order.customer.mktsegment")
db_lineitems.create_index("order.customer.regionname")
db_suppliers.create_index("nation.region")
```