Matemàtica Computacional i Analítica de Dades

Bases de Dades no Relacionals

EL CÒMIC FELIÇ

Informe del Projecte

Marc Bosom - 1606776 Genís Córdoba - 1603639 Laia Escursell - 1600578 Abril Pérez - 1600601

1 Introducció

En aquest projecte es treballarà el disseny, la implementació i la consulta a una base de dades en MongoDB. A partir dels requeriments i les dades que es subministraran, s'ha implementat un script en Python capaç de processar i inserir les dades en una base de dades de MongoDB. Per tal de demostrar que la nostra base de dades funciona correctament, s'ha fet un conjunt de proves: "joc de proves", que són un seguit de queries per tal de demostrar que la base de dades està ben estructurada i dissenyada.

Tots els arxius de Python necessaris i l'arxiu a es troben al https://github.com/genisgui/MongoDB-comics.gitRepositori GitHub. Seguint les instruccions que es troben al fitxer README es poden obtenir i posteriorment executar per tal de visualitzar les diferents queries del joc de proves.

2 Creació de les col·leccions

Per convertir el model Entitat-Relació a un conjunt de col·leccions per a l'ús amb PyMongo, és necessari tenir en compte tant el propi disseny ER com les consultes que es faran posteriorment, perquè aquestes ens indiquen la informació que haurà de ser més accessible i la que no es necessita amb tanta freqüència.

2.1 Selecció de patrons de disseny

El primer que cal fer per transformar el model ER a col·leccions és analitzar les relacions que existeixen entre les diferents entitats i decidir quin dels patrons de conversió a col·leccions s'adecua millor a cada cas.

Relació entre Editorial i Col·lecció

Aquesta és una relació **1-N**, és a dir, que una editorial pot tenir diferents col·leccions però una col·lecció no pot pertànyer a diferents editorials. Això ens deixa amb dues opcions principals: referenciar les diferents col·leccions dins de l'editorial, o encastar tota la informació de les col·leccions dins de l'editorial. Observant el conjunt de consultes que cal fer, veiem que només n'hi ha dues en les que es necessiten les dues entitats i, en ambdós casos, només es necessita el nom de l'editorial, no hi ha cap altre atribut necessari d'aquesta entitat. Per aquest motiu s'ha decidit aplicar el patró referencial: crearem una col·lecció per cada una de les dues entitats, on la col·lecció **Editorial** tindrà per _id el seu nom i l'altra col·lecci'o tindrà un camp que indicarà el nom de l'editorial a la qual pertany.

Relació entre Artista i Publicació

En aquest cas, aquesta relació és **N-N**, un artista pot haver contribuit a diferents publicacions i una publicacio pot comptar amb diferents artistes. Una peculiaritat d'aquesta relació és que està formada per dos relacions diferents, una pels dibuixants i una altra pels guionistes, per tant, serà necessari traspassar aquesta informació a les col·leccions. Per aquest cas, i tenint en compte que les consultes estan relacionades amb el tipus de contribució feta i al nom artístic de l'artista, hem decidit utilitzar una referència estesa. Crearem una col·lecció tant per artista com per publicació, on la primera tindrà per _id el nom artístic i la segona el tindrà de tipus ObjectId. Afegirem a la col·lecció Publicació un camp que serà una llista de subdocuments, on cada un d'aquests conté l'id d'artista (el seu nom) i la contribució que ha fet a la publicació (dibuixant o guionista).

Relació entre Publicació i Personatge

Aquesta també és una relació **N-N**, una publicació pot tenir diferents personatges i un personatge pot pertànyer a diferents publicacions. En aquest cas, hem vist que cada personatge no té gaire informació, només el nom i el tipus, i, a més a més, les consultes relacionades amb aquestes dues entitats demanen informació tant del nom com del tipus de personatge. Per tots aquests motius hem decidit que la millor opció és encastar els personatges dins de cada publicació. Està clar que això provocarà un problema de duplicitat de dades: com que hi ha personatges que pertanyen a divereses publicacions, la seva informació apareixerà diferents vegades dins de la base de dades, però tenint en compte que cada personatge té molt poca informació, creiem que val la pena fer l'encastat.

Relació entre Col·lecció i Publicació

Aquesta relació és de tipus 1-N, és a dir, que una col·lecció pot tenir diverses publicacions però una publicació no pot pertànyer a més d'una col·lecció. Moltes de les consultes demanen molta informació creuada entre col·leccions i publicacions, de manera que en un primer moment haviem decidit encastar les publicacions dins de la seva col·lecció. Tanmateix, al començar a fer la programació ens vam adonar que la gran quantitat de publicacions que podia arribar a tenir una col·lecció i el gran nombre d'atributs que tenen les publicacions feien que cada un dels documents de les col·leccions fós molt extens i complex, de manera que finalment s'ha decidit utilitzar un link. S'ha posat l'ISBN com a __id de la publicació, i cada una de les col·leccions té un camp amb la llista d'ids de les publicacions que conté.

2.2 Estructura de les col·leccions

Finalment, s'han creat quatre col·leccions diferents: Col·lecció, Publicació, Artista i Editorial. L'estructura que tenen aquestes quatre col·leccions són la següent:

Taula 1: Estructura de dades per a una llibreria de còmics

Editorial	
nom	string
responsable	string
país	string
adreça	subdocument: carrer (string), número (int) i població (string)
Col·lecció	
nom	string
exemplars totals	int
editorial	string (_id de l'editorial)
gènere	string
idioma	string
any d'inici	int
any de finalització	int
acabada	bool
publicacions	list of string (_id de les publicacions)
Publicació	
ISBN	string
títol	string
autor	string
nombre de pàgines	int
stock	int
preu	float
personatges	list of subdocuments: nom (string) i tipus (string) de personatge
dibuixants	list of strings (_id de l'artista dibuixant)
guionistes	list of strings (_id de l'artista guionista)
Artista	
nom artístic	string
nom	string
cognoms	string
data de naixement	Date
país	string

3 Joc de Proves

Ara que ja hem creat les col·leccions, farem diferents consultes per comprovar el correcte funcionament de la nostra estructura. A continuació es detallaran les consultes que s'han fet i els resultats obtinguts per cada una d'elles.

Les 5 publicacions amb major preu. Mostrar només el títol i preu.



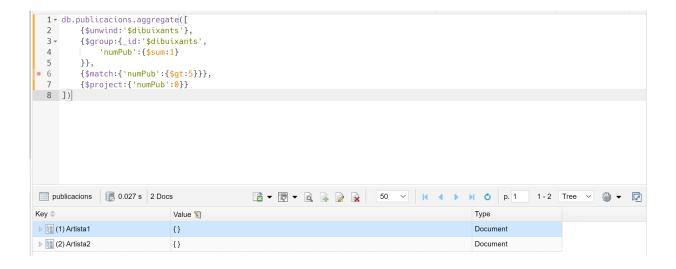
Problema 2

Valor màxim, mínim i mitjà del preus de les publicacions de l'editorial Juniper Books.

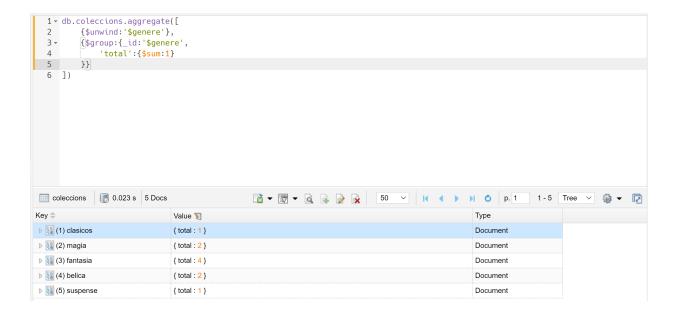


Problema 3

Artistes (nom artístic) que participen en més de 5 publicacions com a dibuixant.

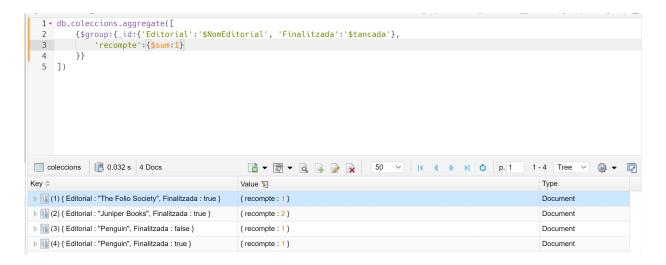


Numero de col·leccions per gènere. Mostra gènere i número total.



Problema 5

Per cada editorial, mostrar el recompte de col·leccions finalitzades i no finalitzades.

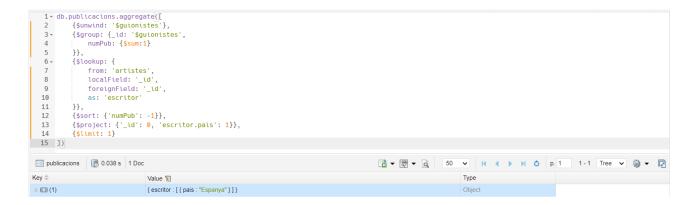


Problema 6

Mostrar les 2 col·leccions ja finalitzades amb més publicacions. Mostrar editorial i nom col·lecció.

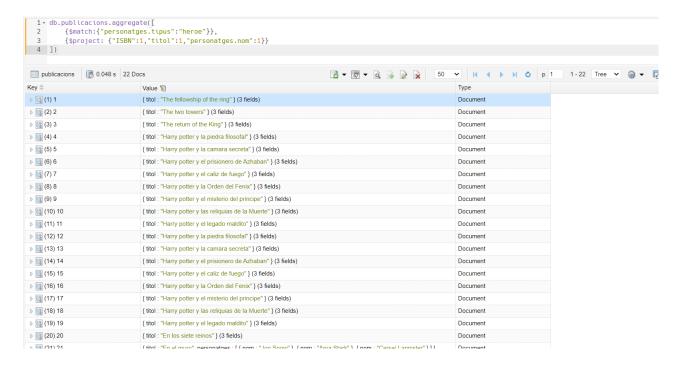
```
1 - db.coleccions.aggregate([
            {$lookup:{
                from:'publicacions',
   4
                localField:'exemplars',
   5
                foreignField:'_id',
   6
                as: 'publicacions'
           }}.
            {$match:{'tancada':true}},
   8
            {\$addFields:\{\numPublicacions\':\{\$size:\\$publicacions\'\}\}, \{\$sort:\{\numPublicacions\':-1\}\},
  10
            {\sproject:{'NomEditorial':1, 'NomColleccio':1, '_id':0}},
  11
  12
  13 ])
                                                                                         coleccions 0.032 s 2 Docs
                                                                                                                            1 - 2 Tree
Key $
                                                                                                                              Type
                                                   Value 🫐
▶ €3 (1)
                                                   { NomColleccio : "Harry Potter", NomEditorial : "Penguin" }
                                                                                                                              Object
▶ €3 (2)
                                                   { NomColleccio : "Harry Potter", NomEditorial : "Juniper Books" }
```

Mostrar el país d'origen de l'artista o artistes que han fet més guions.



Problema 8

Mostrar les publicacions amb tots els personatges de tipus "heroe". Volem destacar que hem entès que hem d'imprimir per pantalla el títol de les publicacions amb tots els seus personatges que siguin del tipus heroi. Per tant, primer hem imprès el títol i, a continuació, la llista dels personatges que compleixin la característica especificada.

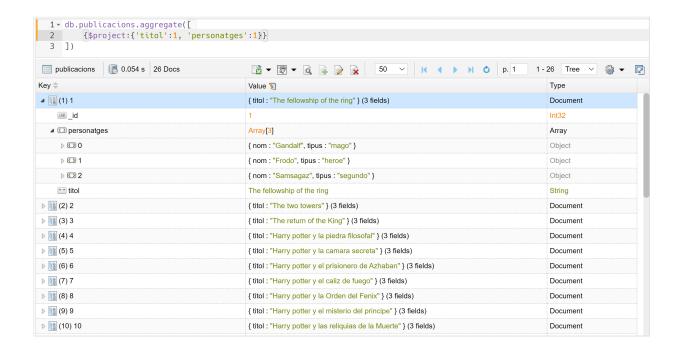


Modificar el preu de les publicacions amb stock superior a 20 exemplars i incrementar-lo un 25%.



Problema 10

Mostrar ISBN i títol de les publicacions conjuntament amb tota la seva informació dels personatges.



4 Distribució de la càrrega de treball

Inicialment vam organitzar una reunió on vam decidir quins patrons de disseny utilitzariem i ens vam distribuir les tasques. De forma general, el Marc Bosom i el Genís Córdoba s'han orientat més a la creació de l'script de python, la Laia Escursell a les consultes, i l'Abril Pérez a la redacció de l'informe, tot i que a la pràctica hi ha moltes tasquest que s'han fet conjuntament encara que només una persona hagi fet el commit de la feina. Quan aquesta s'ha fet conjuntament s'ha especificat a la descripció del commit qui hi ha col·laborat encara que aquesta persona no hagi fet el commit a GitHub.

5 Conclusions

Aquesta pràctica ens ha sigut molt últil per aprendre a utilitzar de forma paralela el mongodb amb python i el NoSQLBooster. Fins ara no haviem utilitzat de forma conjunta aquestes dues eines, i hem vist que és una molt bona opció crear les col·leccions a través de python i després escriure les queries al NoSQLBooster, que permet una execució molt pràctica i més visual.

També hem vist que la creació de les col·leccions a partir de documents excel és més complicada del que sembla a simple vista, perquè cal tenir en compte molts aspectes, com el tipus de dades, els documents duplicats, les relacions entre col·leccions, etc.