

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

Abril 2023

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.git> Repositori 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'adequa 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 · lecció 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 contribuït a diferents publicacions i una publicació 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 duplictat de dades: com que hi ha personatges que pertanyen a diverses 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 havíem 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.

Problema 1

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

```
1 db.publicacions.find({}, {titol:1, preu:1, _id:0}).sort({preu:-1}).limit(5)
```

publicacions 0.026 s 5 Docs		
Key	Value	Type
(1)	{ titol: "Dracula", preu : 125.5 }	Object
(2)	{ titol: "Tragedias", preu : 85.4 }	Object
(3)	{ titol: "Romances", preu : 72.4 }	Object
(4)	{ titol: "Crimen y castigo", preu : 59.4 }	Object
(5)	{ titol: "En el Este", preu : 43.5 }	Object

Problema 2

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

```
1 db.coleccions.aggregate([
2   {$lookup:{
3     from:'publicacions',
4     localField:'exemplars',
5     foreignField:'_id',
6     as:'publicacions'
7   }},
8   {$unwind:'$publicacions'},
9   {$match:{'NomEditorial':'Juniper Books'}},
10  {$group:{'_id':'$NomEditorial',
11            'preuMax':{'$max':'$publicacions.preu'},
12            'preuMin':{'$min':'$publicacions.preu'},
13            'preuAvg':{'$avg':'$publicacions.preu'}
14          }}
15 ])
```

coleccions 0.026 s 1 Doc		
Key	Value	Type
(1) Juniper Books	{ preuMax : 32.5, preuMin : 27.85, preuAvg : 29.11818181818182 } (4 fields)	Document

Problema 3

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

```
1 db.publicacions.aggregate([
2   {$unwind:'$dibuixants'},
3   {$group:{'_id':'$dibuixants',
4             'numPub':{'$sum':1}}},
5   {$match:{'numPub':{'$gt':5}}},
6   {$project:{'numPub':0}}
7 ])
```

publicacions 0.027 s 2 Docs		
Key	Value	Type
(1) Artista1	{}	Document
(2) Artista2	{}	Document

Problema 4

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

```
1 db.coleccions.aggregate([
2   {$unwind:'$genere'},
3   {$group:{_id:'$genere',
4             'total':{$sum:1}}
5   })
6 ])
```

coleccions 0.023 s 5 Docs

Key	Value	Type
(1) clasicos	{ total : 1 }	Document
(2) magia	{ total : 2 }	Document
(3) fantasia	{ total : 4 }	Document
(4) belica	{ total : 2 }	Document
(5) suspense	{ total : 1 }	Document

Problema 5

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

```
1 db.coleccions.aggregate([
2   {$group:{_id:{'Editorial': '$NomEditorial', 'Finalitzada': '$tancada'},
3             'recompte':{$sum:1}}
4   })
5 ])
```

coleccions 0.032 s 4 Docs

Key	Value	Type
(1) { Editorial : "The Folio Society", Finalitzada : true }	{ recompte : 1 }	Document
(2) { Editorial : "Juniper Books", Finalitzada : true }	{ recompte : 2 }	Document
(3) { Editorial : "Penguin", Finalitzada : false }	{ recompte : 1 }	Document
(4) { Editorial : "Penguin", Finalitzada : true }	{ recompte : 1 }	Document

Problema 6

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

```
1 db.coleccions.aggregate([
2   {$lookup:{
3     from:'publicacions',
4     localField:'exemplars',
5     foreignField:'_id',
6     as:'publicacions'
7   }},
8   {$match:{'tancada':true}},
9   {$addFields:{'numPublicacions':{'$size':'$publicacions'}}},
10  {$sort:{'numPublicacions':-1}},
11  {$project:{'NomEditorial':1, 'NomCol·leccio':1, '_id':0}},
12  {$limit:2}
13 ])
```

coleccions0.032 s2 Docs

50

p. 11 - 2Tree

Key	Value	Type
(1)	{ NomCol·leccio : "Harry Potter", NomEditorial : "Penguin" }	Object
(2)	{ NomCol·leccio : "Harry Potter", NomEditorial : "Juniper Books" }	Object

Problema 7

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

```
1- db.coleccions.aggregate([
2-   {$lookup:{
3-     from:'publicacions',
4-     localField:'exemplars',
5-     foreignField:'_id',
6-     as:'publicacions'
7-   }},
8-   {$unwind:'$publicacions'},
9-   {$match: {'NomEditorial': 'Juniper Books'}},
10-  {$group: { _id: '$NomEditorial',
11-    'preuMax': { $max: '$publicacions.preu' },
12-    'preuMin': { $min: '$publicacions.preu' },
13-    'preuAvg': { $avg: '$publicacions.preu' }
14-  }}
15- ])
```

coleccions

0.026 s

1 Doc

50

p. 1

1 - 1

Tree

Key	Value	Type
(1) Juniper Books	{ preuMax: 32.5, preuMin: 27.85, preuAvg: 29.118181818182 } (4 fields)	Document

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.

```

1 db.publicacions.aggregate([
2   {$match: {"personatges.tipus": "heroe"}},
3   {$project: {"ISBN":1,"titol":1,"personatges.nom":1}}
4 ])
```

publicacions

0.048 s

22 Docs

50

p. 1

1 - 22

Tree

Key	Value	Type
▶ (1) 1	{ titol : "The fellowship of the ring" } (3 fields)	Document
▶ (2) 2	{ titol : "The two towers" } (3 fields)	Document
▶ (3) 3	{ titol : "The return of the King" } (3 fields)	Document
▶ (4) 4	{ titol : "Harry potter y la piedra filosofal" } (3 fields)	Document
▶ (5) 5	{ titol : "Harry potter y la camara secreta" } (3 fields)	Document
▶ (6) 6	{ titol : "Harry potter y el prisionero de Azhaban" } (3 fields)	Document
▶ (7) 7	{ titol : "Harry potter y el caliz de fuego" } (3 fields)	Document
▶ (8) 8	{ titol : "Harry potter y la Orden del Fenix" } (3 fields)	Document
▶ (9) 9	{ titol : "Harry potter y el misterio del principe" } (3 fields)	Document
▶ (10) 10	{ titol : "Harry potter y las reliquias de la Muerte" } (3 fields)	Document
▶ (11) 11	{ titol : "Harry potter y el legado maldito" } (3 fields)	Document
▶ (12) 12	{ titol : "Harry potter y la piedra filosofal" } (3 fields)	Document
▶ (13) 13	{ titol : "Harry potter y la camara secreta" } (3 fields)	Document
▶ (14) 14	{ titol : "Harry potter y el prisionero de Azhaban" } (3 fields)	Document
▶ (15) 15	{ titol : "Harry potter y el caliz de fuego" } (3 fields)	Document
▶ (16) 16	{ titol : "Harry potter y la Orden del Fenix" } (3 fields)	Document
▶ (17) 17	{ titol : "Harry potter y el misterio del principe" } (3 fields)	Document
▶ (18) 18	{ titol : "Harry potter y las reliquias de la Muerte" } (3 fields)	Document
▶ (19) 19	{ titol : "Harry potter y el legado maldito" } (3 fields)	Document
▶ (20) 20	{ titol : "En los siete reinos" } (3 fields)	Document
▶ (21) 21	{ titol : "En el mundo de los heroes" } (3 fields)	Document

Problema 9

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

```
1 db.publicacions.updateMany([
2   {'stock':{'$gt':20}},
3   {'$set':{'preu':{'$multiply':['$preu', 1.25]}}}]
4 ])
```

0.026 s

Key	Value	Type
(1)	{ acknowledged : true, matchedCount : 7, modifiedCount : 7 }	Object
acknowledged	true	Bool
matchedCount	7	Int32
modifiedCount	7	Int32

Problema 10

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

```
1 db.publicacions.aggregate([
2   {'$project':{'titol':1, 'personatges':1}}
3 ])
```

publicacions 0.054 s 26 Docs

Key	Value	Type
(1) 1	{ titol : "The fellowship of the ring" } (3 fields)	Document
_id	1	Int32
personatges	Array[3]	Array
0	{ nom : "Gandalf", tipus : "mago" }	Object
1	{ nom : "Frodo", tipus : "heroe" }	Object
2	{ nom : "Samsagaz", tipus : "segundo" }	Object
titol	The fellowship of the ring	String
(2) 2	{ titol : "The two towers" } (3 fields)	Document
(3) 3	{ titol : "The return of the King" } (3 fields)	Document
(4) 4	{ titol : "Harry potter y la piedra filosofal" } (3 fields)	Document
(5) 5	{ titol : "Harry potter y la camara secreta" } (3 fields)	Document
(6) 6	{ titol : "Harry potter y el prisionero de Azaban" } (3 fields)	Document
(7) 7	{ titol : "Harry potter y el caliz de fuego" } (3 fields)	Document
(8) 8	{ titol : "Harry potter y la Orden del Fenix" } (3 fields)	Document
(9) 9	{ titol : "Harry potter y el misterio del principe" } (3 fields)	Document
(10) 10	{ titol : "Harry potter y las reliquias de la Muerte" } (3 fields)	Document

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 tasquestes 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 útil per aprendre a utilitzar de forma paral·lela el mongodb amb python i el NoSQLBooster. Fins ara no havíem 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·lents 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.