Tasca S9.01. Consultes amb MongoDB

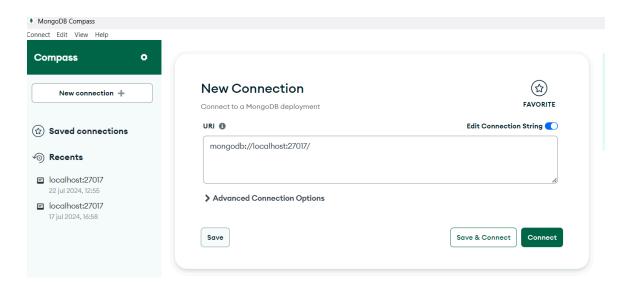
Treballarem amb una base de dades que conté col·leccions relacionades amb una aplicació d'entreteniment cinematogràfic:

- users: Emmagatzema informació d'usuaris/es, incloent-hi noms, emails i contrasenyes xifrades.
- theatres: Conté dades de cinemes, com ID, ubicació (direcció i coordenades geogràfiques).
- sessions: Guarda sessions d'usuari, incloent-hi ID d'usuari i tokens JWT per a l'autenticació.
- o movies: Inclou detalls de pel·lícules, com a trama, gèneres, durada, elenc, comentaris, any de llançament, directors, classificació i premis.
- o comments: Emmagatzema comentaris d'usuaris/es sobre pel·lícules, amb informació de l'autor/a del comentari, ID de la pel·lícula, text del comentari i la data.

Duràs a terme algunes consultes que et demana el client/a, el qual està mesurant si seràs capaç o no de fer-te càrrec de la part analítica del projecte vinculat amb la seva base de dades.

Nivell 1

Crea una base de dades amb MongoDB utilitzant com a col·leccions els arxius adjunts.



Primer de tot, creem una base de dades amb el nom de **test** i triem una primera col·lecció, per exemple "comments". Una vegada creada la base de dades i importat el primer fitxer JSON, fem la importació de la resta d'arxius. Com que mai he treballat

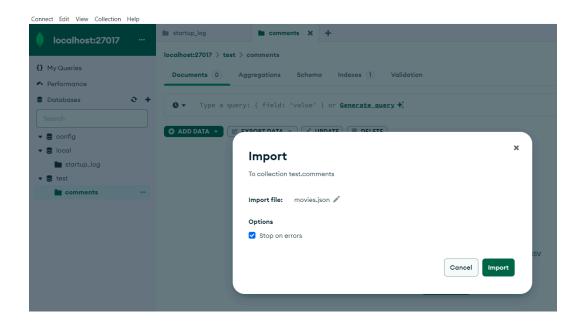
amb arxius JSON abans obro els arxius des del Visual Studio per explorar una mica les dades:



Observo que, com esperàvem, el format és vàlid per fer una importació a *MongoDB Compass*, ja que els arxius estan separats per comes en una matriu.

Així doncs, ja podem fer els passos per fer la importació de la resta d'arxius JSON en una col·lecció:

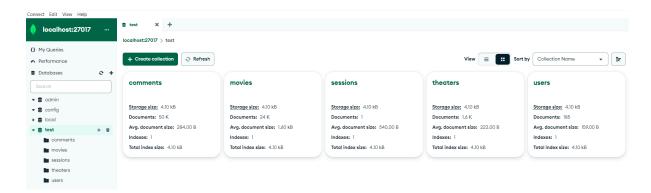
- 1. Seleccionada la base de dades **test** cliquem *ADD DATA* i seleccionem *Import JSON o CSV file*.
- 2. Importem "movies.json", marcant l'opció perquè es pari en cas d'error:



3. Ens informa que el procés s'ha fet correctament, per la qual cosa ara anem al símbol "+" que hi ha al costat de la base de dades **test** i creem la resta de col·leccions:



4. Hem carregat tots els arxius JSON en cinc col·leccions:

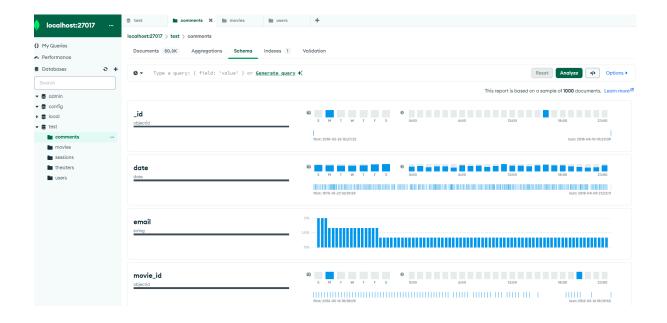


Test és la nostra bd, i les col·leccions són: comments, movies, sessions, theaters i users. Per fer un símil amb SQL, les col·leccions serien com l'equivalent a les taules.

Abans de passar a fer les consultes, explorem de manera ràpida l'esquema de cada col·lecció mitjançant l'opció d'anàlisi de l'esquema que ens ofereix informació sobre la freqüència, tipus i els intervals de camps del conjunt de dades.

Cal recordar que, gràcies a aquest modelatge flexible de dades de MongoDB, hem pogut crear les col·leccions a partir dels documents amb diferents estructures.

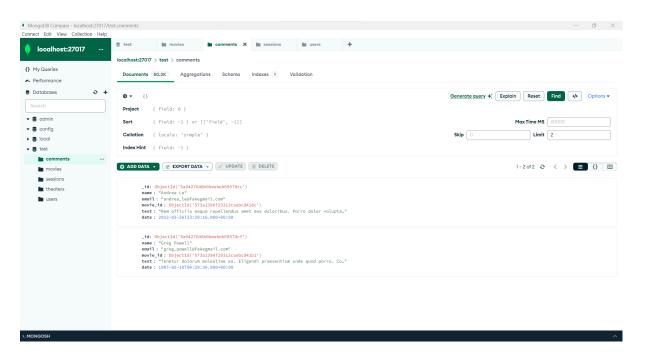
A continuació una mostra de l'esquema de "comments" que es genera a partir d'una mostra de 1.000 documents, que serien com els registres. Ens permet veure un resum de les dades d'una col·lecció.



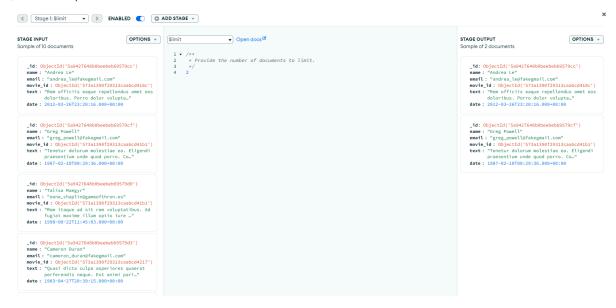
- Exercici 1

1. Mostra els 2 primers comentaris que hi ha en la base de dades.

Una manera és utilitzant la interfície de *Compass*, des de la col·lecció "comments", la pestanya documents, fer clic a les Opcions i limitar a 2:



També podem trobar els dos primers resultats amb Aggregations --> Add Stage --> \$limit --> posar el 2 a number:



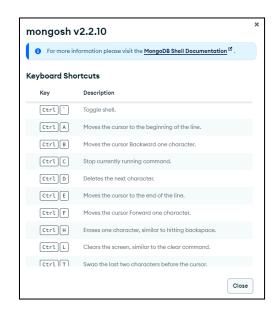
I en últim lloc podem obtenir els mateixos resultats utilitzant el shell incrustat mongosh de la següent manera:

```
MongoDB Compass - localhost2/2017 ---

localhost27017 ---

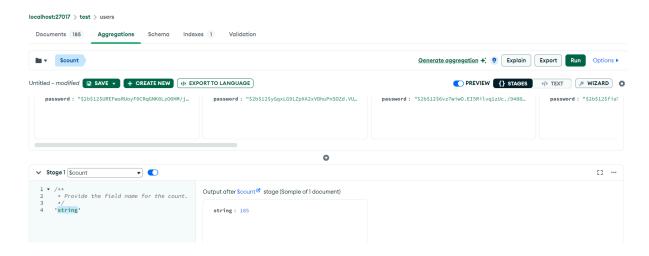
test | movies | test | localhost27017 | movies | mov
```

Aquí ens faciliten els *shortcuts* per agilitzar el procés, i un enllaç per accedir a la Documentació:



2. Quants usuaris tenim registrats?

Fent clic a *Aggregations* de la col·lecció *users* ja ens informa que hi ha 185 documents que equival a usuaris registrats, però repetim els mateixos passos des d'*Aggregations* --> Add Stage --> \$count :

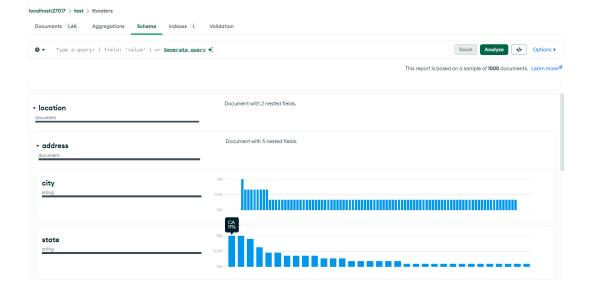


Amb mongosh obtenim el mateix resultat de **185 usuaris registrats**, utilitzant la funció "countDocuments()":



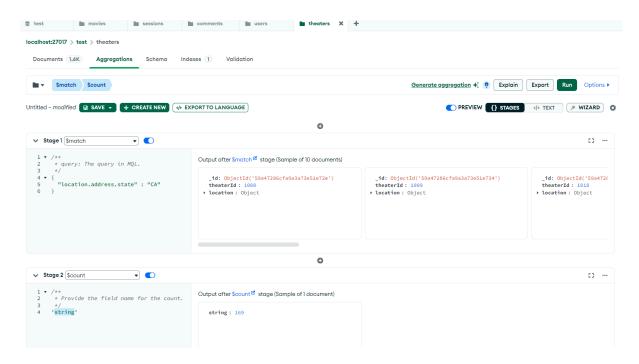
3. Quants cinemes hi ha en l'estat de Califòrnia?

Primer he anat a l'esquema i he vist que els documents tenen una estructura anidada. Aquesta informació la trobem de manera que el camp 'state' està anidat dins 'address' i finalment dins de 'location'. Califòrnia es representa com a CA i és de tipus cadena.

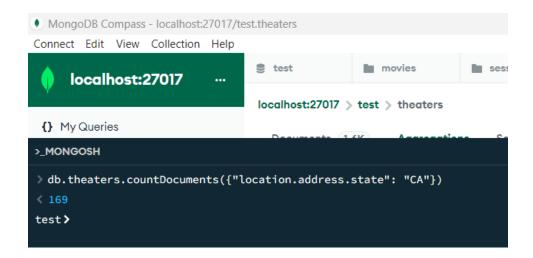


Així doncs des de la interfície anem a *theaters* --> Stage 1 **\$match** i substituïm *query* per: "**location.address.state" : "CA"**

Per accedir a camps anidats utilitzem el . i com a filtre d'igualtat que busqui el valor "CA" perquè coincideixi en la nostra consulta. Amb això l'output ens retorna 10 documents, però encara ens falta un últim pas, que és fer el recompte de cinemes amb un Stage 2 **\$count**. El resultat ara sí, és de 169:

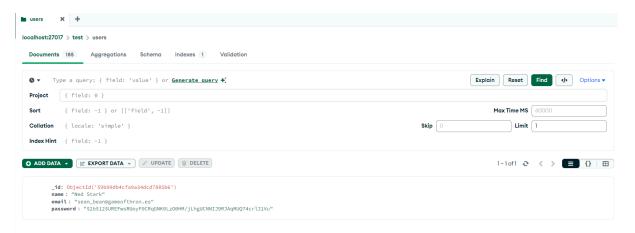


Ara utilitzant mongosh també ens mostra que Califòrnia té 169 cinemes:

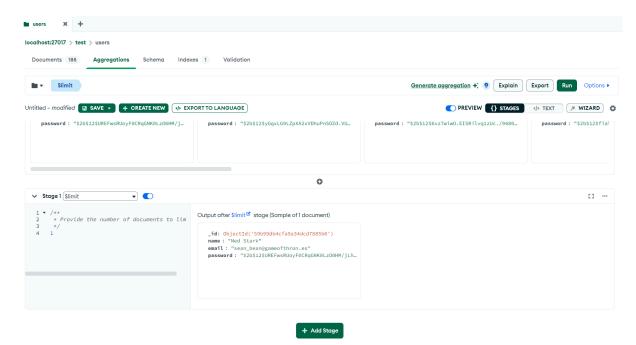


4. Quin va ser el primer usuari/ària en registrar-se?

Des de *Compass* col·lecció "users", pestanya de documents, fer clic a les Opcions i limitar a 1:



També així:



I per acabar amb mongosh, el resultat sempre és l'usuari amb el nom **Ned Stark**:

```
>_MONGOSH

> db.users.findOne()

< {
    __id: ObjectId('59b99db4cfa9a34dcd7885b6'),
    name: 'Ned Stark',
    email: 'sean_bean@gameofthron.es',
    password: '$2b$12$UREFwsRUoyF0CRqGNK0LzO0HM/jLhgUCNNIJ9RJAqMUQ74crlJ1Vu'
    }
test>
```

5. Quantes pel·lícules de comèdia hi ha en la nostra base de dades?

Anem a la col·lecció "movies" des de l'esquema veiem que type és un *string* amb dos valors: "movie" i "series":

```
type
string movie
```

Amb la consola és més ràpid:

```
>_MONGOSH

> db.movies.distinct("type")

< [ 'movie', 'series' ]

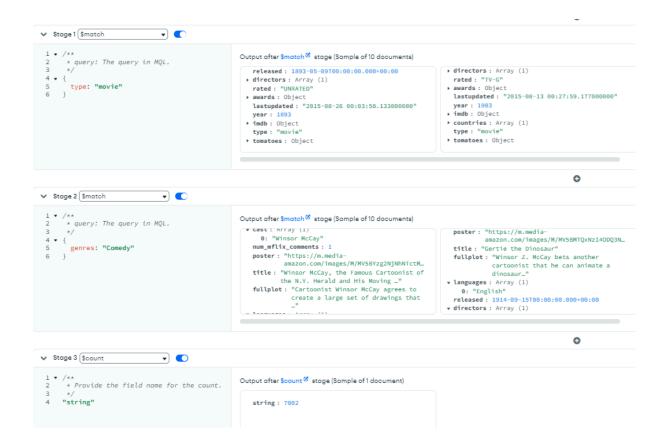
test>
```

Abans també vull veure la classificació de gèneres existent, torno a utilitzar la consola per veure com s'anomenen els gèneres:

```
>_MONGOSH
> db.movies.distinct("genres")
   'Action',
                'Adventure', 'Animation',
   'Biography', 'Comedy', 'Crime',
   'Documentary', 'Drama',
                            'Family',
   'Fantasy',
                'Film-Noir', 'History',
   'Mystery',
                'News',
                            'Romance',
                 'Thriller', 'War',
   'Talk-Show',
   'Western'
test>
```

Veiem, doncs, que hem de comptar el total de pel·lícules on gènere és "Comedy".

Una manera d'obtenir el resultat és des d'Aggregations amb aquestes tres fases:



El resultat són 7002 pel·lícules de comèdia a la base de dades.

També ho podem obtenir així:

```
{} My Queries

>_MONGOSH

> db.movies.countDocuments({"$and": [ {"type": "movie"}, {"genres": "Comedy"}]})

< 7002
test>
```

- Exercici 2

Mostra'm tots els documents de les pel·lícules produïdes en 1932, però que el gènere sigui drama o estiguin en francès.

Ens demanen tots els documents de la col·lecció "movies" que compleixin aquestes dues condicions:

- 1. Les pel·lícules han de ser produïdes en 1932 (ho trobem a {"type": "movie"}, {"year": 1932).
- 2. El gènere ha de ser drama (ho trobem a {"genres": "Drama"} l'idioma ha de ser francès (ho trobem a {"languages": "French}.

Consideració!: Tot i veure que a **db.movies.distinct("languages")** ens apareix també el francès amb llengua dels signes, en aquest cas no l'inclourem.

```
'Finnish',
'Flemish',
'French',
'French Sign Language',
```

Obtenim els resultats amb Compass:

Stage 1 \$match

{type: "movie"}

```
V Stagel Smatch

Output offer Smotch stage (Sample of 10 documents)

* * query: The query in MQL.

* * query: The query in MQL.

* * query: The query in MQL.

* * (10 bjectId('573a1399f29313caabcd426*)

* plot: "A group of bandits stage as brazen train hold-up, only to find a determ."

* penres: Array (1)

* runtine: 1

* cast: Array (2)

* num_mflix_comments: 1

* title: "Blacksmith Scene"

fullplot: "A stationary camera looks at a

* title: "The Great Train Robbery"

* poster: "MrUSHMTUSHjESNZYY...

* title: "The Great Train Robbery"

* poster: "MrUSHMTUSHjESNZYY...

* title: "The Great Train Robbery"
```

Stage 2 \$match

{year: 1932}

Stage 3 \$match

{\$or: [{ genres: "Drama" }, { languages: "French" }]}

Stage 4 \$count

De la consulta n'extreiem que tenim 18 documents amb aquestes condicions.

```
V Stage 4 Socurit

1 * /**
2 * Provide the field name for the count.
3 */
4 'string'

Output after Socurit® stage (Sample of I document)

string: 18
```

Amb la consola tots els resultats es veuen aquí:

I si fem un countDocuments obtenim un total de 18 documents.

- Exercici 3

Mostra'm tots els documents de pel·lícules estatunidenques que tinguin entre 5 i 9 premis que van ser produïdes entre 2012 i 2014.

Per aquest cas ens demanen tots els documents de la col·lecció "movies" que compleixin aquests requisits:

- 1. Pel·lícules estatunidenques: El camp "type" ha de ser "movie" i "countries" ha de contenir el valor "USA".
- 2. Entre 5 i 9 premis: El camp "awards.wins" ha de contenir un nombre de premis entre 5 i 9.
- 3. Produïdes entre 2012 i 2014: El camp "year" ha de ser entre 2012 i 2014.

Amb Compass des de la visualització </> text l'output final ens dona 162 documents:

Amb la consola fem:

```
>_MONGOSH

> db.movies.countDocuments({
    "type": "movie",
    "countries": "USA",
    "awards.wins": { $gte: 5, $lte: 9 },
    "year": { $gte: 2012, $lte: 2014 }
    })
< 162
test>
```

Alguns títols d'aquestes pel·lícules són: 'The Secret Life of Walter Mitty', 'The Croods' o 'The Book Thief' com es pot veure aquí:

```
> db.movies.find(
     "type": "movie",
     "countries": "USA",
     "awards.wins": { $gte: 5, $lte: 9 },
     "year": { $gte: 2012, $lte: 2014 }
    title: 1
 ).limit(15)
< €
   title: 'The Secret Life of Walter Mitty'
   title: 'The Croods'
   _id: ObjectId('573a13b9f29313caabd4dd15'),
   _id: ObjectId('573a13b9f29313caabd4de39'),
   title: 'World War Z'
   _id: ObjectId('573a13b9f29313caabd4f21f'),
   _id: ObjectId('573a13bbf29313caabd5271d'),
   title: 'Mr Hockey: The Gordie Howe Story'
   _id: ObjectId('573a13bcf29313caabd57341'),
```

Nivell 2

- Exercici 1

Compte quants comentaris escriu un usuari/ària que utilitza "GAMEOFTHRON.ES" com a domini de correu electrònic.

Fem la consulta directament a la consola, l'operador **\$regex** permet buscar els documents basats en patrons de text mitjançant expressions regulars. Permet cercar coincidències en camps de text, en el nostre cas escrivint **@gameofthron.es\$** tenim en compte que el correu ha d'acabar d'aquesta manera.

```
>_MONGOSH

> db.comments.countDocuments({
    "email": { $regex: /@gameofthron.es$/}
})
< 22841
test>
```

El resultat és que hi ha 22.841 comentaris d'aquest/a usuari/ària amb aquest criteri.

- Exercici 2

Quants cinemes hi ha en cada codi postal situats dins de l'estat Washington D. C. (DC)?

Fet directament per consola el resultat ens dona que hi ha 3 cinemes. Hem filtrat els documents per incloure només els teatres ubicats a l'estat "DC" i grupat pel codi postal (zipcode). En últim lloc, comptem quants teatres hi ha a cada codi postal:

Nivell 3

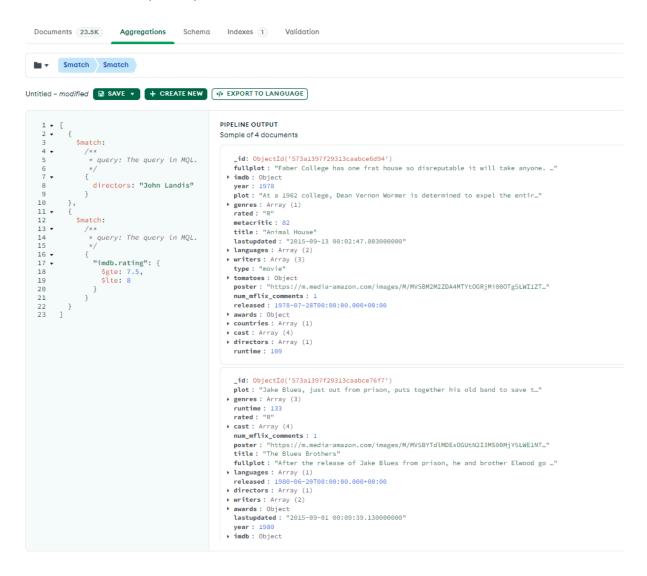
- Exercici 1

Troba totes les pel·lícules dirigides per John Landis amb una puntuació IMDb (Internet Movie Database) d'entre 7,5 i 8.

Trobem 4 pel·lícules que compleixen aquests criteris. També mostrem els títols d'aquestes pel·lícules :

```
>_MONGOSH
                                                          > db.movies.find(
                                                               directors: "John Landis".
                                                               "imdb.rating": { $gte: 7.5, $lte: 8 }
                                                             { title: 1, _id: 0 }
>_MONGOSH
                                                           ).pretty()
> db.movies.find(
                                                             title: 'Animal House'
    {$and:[
     {directors:"John Landis"},
                                                             title: 'The Blues Brothers'
     {"imdb.rating":{$gte:7.5, $lte:8}}
  ).count()
test>
                                                          test>
```

Amb *Compass* tenim exactament el mateix resultat i també podem veure tota la informació de les quatre pel·lícules:

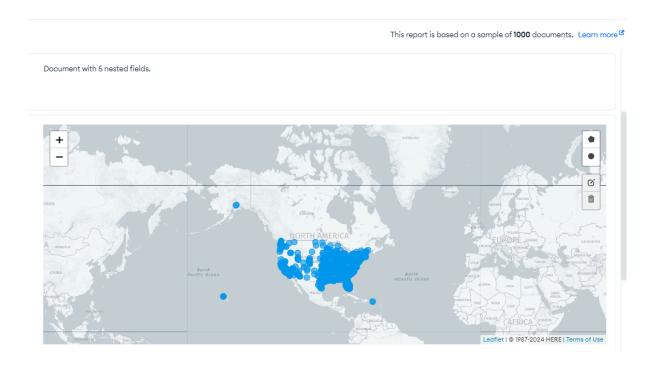


- Exercici 2

Mostra en un mapa la ubicació de tots els teatres de la base de dades.

Per mostrar el mapa, hem d'anar a la col·lecció **theaters** i si mirem l'esquema veiem que dins de **location** trobem **geo** que mostra un mapa, perquè és de tipus coordenades. Alhora dins hi té **coordinates** que mostra el mateix mapa.

Faig una captura d'aquesta mapa on es mostra la ubicació d'una part dels teatres, ja que si llegim l'avís de la part superior dreta, diu que només veiem una mostra basada en 1000 documents.



Des d'*Aggregations* si fem un recompte, surt que tenim 1564 teatres, per la qual cosa no es pinten 564 teatres. Tan sols estaríem mostrant un 64% del total.



Opino que amb aquesta mostra del mapa d'ubicacions, és prou rellevant per tenir una idea de la distribució dels teatres de la base de dades.