

Práctica con el dataset sample_mflix de MongoDB

Bienvenido a esta práctica en la que exploraremos el dataset **sample_mflix**, un conjunto de datos relacionado con películas incluido en los datasets de muestra de MongoDB. A través de esta práctica, adquirirás experiencia trabajando con consultas y operaciones de agregación en MongoDB.

Ejercicios propuestos:

1. Número de películas lanzadas cada año

Haz una consulta que muestre cuántas películas se lanzaron en cada año.

```
> db.movies.aggregate([
  { $group: { _id: { $year: "$released" }, count: { $sum: 1 } } },
  { $sort: { _id: 1 } }
])
< {
  _id: null,
  count: 471
}
{
  _id: 1903,
  count: 1
}
{
  _id: 1909,
  count: 1
}
{
  _id: 1911,
  count: 2
}
{
  _id: 1913,
  count: 1
}
{
  _id: 1914,
  count: 3
}
{
  _id: 1915,
  count: 2
}
}
```

```
{
  _id: 1916,
  count: 2
}
{
  _id: 1917,
  count: 2
}
{
  _id: 1918,
  count: 1
}
{
  _id: 1919,
  count: 1
}
{
  _id: 1920,
  count: 4
}
{
  _id: 1921,
  count: 3
}
{
  _id: 1922,
  count: 2
}
{
  _id: 1923,
  count: 4
}
{
  _id: 1924,
  count: 6
}
```

```
{
  _id: 1925,
  count: 3
}
{
  _id: 1926,
  count: 6
}
{
  _id: 1927,
  count: 2
}
{
  _id: 1928,
  count: 5
}
Type "it" for more
> it
< {
  _id: 1929,
  count: 8
}
{
  _id: 1930,
  count: 14
}
{
  _id: 1931,
  count: 18
}
{
  _id: 1932,
  count: 17
}
```

```
{
  _id: 1933,
  count: 19
}
{
  _id: 1934,
  count: 19
}
{
  _id: 1935,
  count: 30
}
{
  _id: 1936,
  count: 27
}
{
  _id: 1937,
  count: 27
}
{
  _id: 1938,
  count: 42
}
{
  _id: 1939,
  count: 18
}
{
  _id: 1940,
  count: 26
}
{
  _id: 1941,
  count: 20
}
```

```
{
  _id: 1942,
  count: 25
}
{
  _id: 1943,
  count: 26
}
{
  _id: 1944,
  count: 20
}
{
  _id: 1945,
  count: 31
}
{
  _id: 1946,
  count: 28
}
{
  _id: 1947,
  count: 28
}
{
  _id: 1948,
  count: 52
}
Type "it" for more
> it
< {
  _id: 1949,
  count: 45
}
```

```
{
  _id: 1950,
  count: 61
}
{
  _id: 1951,
  count: 42
}
{
  _id: 1952,
  count: 41
}
{
  _id: 1953,
  count: 56
}
{
  _id: 1954,
  count: 50
}
{
  _id: 1955,
  count: 55
}
{
  _id: 1956,
  count: 61
}
{
  _id: 1957,
  count: 65
}
{
  _id: 1958,
  count: 60
}
```

```

{
  _id: 1959,
  count: 66
}
{
  _id: 1960,
  count: 72
}
{
  _id: 1961,
  count: 67
}
{
  _id: 1962,
  count: 75
}
{
  _id: 1963,
  count: 59
}
{
  _id: 1964,
  count: 75
}
{
  _id: 1965,
  count: 75
}
{
  _id: 1966,
  count: 84
}
{
  _id: 1967,
  count: 63
}

```

```

{
  _id: 1968,
  count: 96
}
Type "it" for more
> it
< {
  _id: 1969,
  count: 81
}
{
  _id: 1970,
  count: 121
}
{
  _id: 1971,
  count: 103
}
{
  _id: 1972,
  count: 105
}
{
  _id: 1973,
  count: 103
}
{
  _id: 1974,
  count: 97
}
{
  _id: 1975,
  count: 104
}

```

```

{
  _id: 1976,
  count: 99
}
{
  _id: 1977,
  count: 123
}
{
  _id: 1978,
  count: 124
}
{
  _id: 1979,
  count: 148
}
{
  _id: 1980,
  count: 162
}
{
  _id: 1981,
  count: 156
}
{
  _id: 1982,
  count: 180
}
{
  _id: 1983,
  count: 151
}
{
  _id: 1984,
  count: 176
}

```

```

{
  _id: 1985,
  count: 201
}
{
  _id: 1986,
  count: 203
}
{
  _id: 1987,
  count: 198
}
{
  _id: 1988,
  count: 233
}
Type "it" for more
> it
< {
  _id: 1989,
  count: 221
}
{
  _id: 1990,
  count: 217
}
{
  _id: 1991,
  count: 245
}
{
  _id: 1992,
  count: 234
}

```

```

{
  _id: 1993,
  count: 263
}
{
  _id: 1994,
  count: 262
}
{
  _id: 1995,
  count: 334
}
{
  _id: 1996,
  count: 385
}
{
  _id: 1997,
  count: 437
}
{
  _id: 1998,
  count: 494
}
{
  _id: 1999,
  count: 508
}
{
  _id: 2000,
  count: 555
}
{
  _id: 2001,
  count: 588
}
{
  _id: 2002,
  count: 625
}
{
  _id: 2003,
  count: 615
}
{
  _id: 2004,
  count: 645
}
{
  _id: 2005,
  count: 689
}
{
  _id: 2006,
  count: 736
}
{
  _id: 2007,
  count: 800
}
{
  _id: 2008,
  count: 871
}
Type "it" for more
it
{
  _id: 2009,
  count: 857
}
{
  _id: 2010,
  count: 868
}
{
  _id: 2011,
  count: 909
}
{
  _id: 2012,
  count: 903
}
{
  _id: 2013,
  count: 1063
}
{
  _id: 2014,
  count: 1089
}
{
  _id: 2015,
  count: 773
}
{
  _id: 2016,
  count: 10
}

```

2. Calificación promedio de IMDb por género

Encuentra la calificación promedio de IMDb para películas en cada género.

```
> db.movies.aggregate([
  { $unwind: "$genres" },
  { $group: { _id: "$genres", avg_rating: { $avg: "$imdb.rating" } } },
  { $sort: { avg_rating: -1 } }
])
< {
  _id: 'Film-Noir',
  avg_rating: 7.397402597402598
}
{
  _id: 'Short',
  avg_rating: 7.377574370709382
}
{
  _id: 'Documentary',
  avg_rating: 7.365679824561403
}
{
  _id: 'News',
  avg_rating: 7.252272727272728
}
{
  _id: 'History',
  avg_rating: 7.1696100917431185
}
{
  _id: 'War',
  avg_rating: 7.128591954022989
}
{
  _id: 'Biography',
  avg_rating: 7.087984189723319
}
```

```
{
  _id: 'Talk-Show',
  avg_rating: 7
}
{
  _id: 'Animation',
  avg_rating: 6.89669603524229
}
{
  _id: 'Music',
  avg_rating: 6.883333333333334
}
{
  _id: 'Western',
  avg_rating: 6.823553719008264
}
{
  _id: 'Drama',
  avg_rating: 6.803377338624768
}
{
  _id: 'Sport',
  avg_rating: 6.749041095890411
}
{
  _id: 'Crime',
  avg_rating: 6.688585405625764
}
{
  _id: 'Musical',
  avg_rating: 6.665831435079727
}
{
  _id: 'Romance',
  avg_rating: 6.6564272782136396
}
```

```
  _id: 'Mystery',
  avg_rating: 6.527425044091711
}
{
  _id: 'Adventure',
  avg_rating: 6.493680884676145
}
{
  _id: 'Comedy',
  avg_rating: 6.450214658080344
}
{
  _id: 'Fantasy',
  avg_rating: 6.3829847908745245
}
Type "it" for more
> it
< {
  _id: 'Action',
  avg_rating: 6.347098402018503
}
{
  _id: 'Family',
  avg_rating: 6.3296712109061755
}
{
  _id: 'Thriller',
  avg_rating: 6.304498977505112
}
{
  _id: 'Sci-Fi',
  avg_rating: 6.123609653725079
}
{
  _id: 'Horror',
  avg_rating: 5.7847098976:
}
```

Calendar

3. Top 5 películas con más alta calificación en IMDb

Lista las 5 películas con la calificación IMDb más alta.

```
> db.movies.aggregate([
  {
    $addFields: {
      "imdb.rating": {
        $toDouble: {
          $cond: { if: { $eq: [{ $type: "$imdb.rating" }, "string"] }, then: { $cond: { if: { $eq: ["$imdb.rating", ""] }, then: null, else: "$imdb.rating" } }, else: "$imdb.rating" }
        }
      },
      "imdb.votes": {
        $toLong: {
          $cond: { if: { $eq: [{ $type: "$imdb.votes" }, "string"] }, then: { $cond: { if: { $eq: ["$imdb.votes", ""] }, then: null, else: "$imdb.votes" } }, else: "$imdb.votes" }
        }
      }
    }
  },
  { $sort: { "imdb.rating": -1 } },
  { $limit: 5 },
  { $project: { title: 1, "imdb.rating": 1, "imdb.votes": 1 } }
])
{
  "_id": ObjectId('573a139ff29313caabd083c4'),
  "title": 'Band of Brothers',
  "imdb": {
    "rating": 9.6,
    "votes": 183802
  }
}
{
  "_id": ObjectId('573a13b8f29313caabd4c241'),
  "title": 'Planet Earth',
  "imdb": {
    "rating": 9.5,
    "votes": 82896
  }
}
```

```
{
  _id: ObjectId('573a13f0f29313caabdda542'),
  title: 'A Brave Heart: The Lizzie Velasquez Story',
  imdb: {
    rating: 9.4,
    votes: 45
  }
}
{
  _id: ObjectId('573a13a3f29313caabd0e77b'),
  title: 'The Civil War',
  imdb: {
    rating: 9.4,
    votes: 4624
  }
}
{
  _id: ObjectId('573a1398f29313caabcebc0b'),
  title: 'The Civil War',
  imdb: {
    rating: 9.4,
    votes: 4625
  }
}
```

4. Número total de películas y duración promedio por director

Determina el número total de películas y la duración promedio para cada director.

```
> db.movies.aggregate([
  { $unwind: "$directors" },
  { $group: {
    _id: "$directors",
    total_movies: { $sum: 1 },
    avg_duration: { $avg: "$runtime" }
  }
},
{ $sort: { total_movies: -1 } }
])
< {
  _id: 'Woody Allen',
  total_movies: 40,
  avg_duration: 96.075
}
{
  _id: 'Martin Scorsese',
  total_movies: 32,
  avg_duration: 132.84375
}
{
  _id: 'Takashi Miike',
  total_movies: 31,
  avg_duration: 116.96774193548387
}
{
  _id: 'Sidney Lumet',
  total_movies: 29,
  avg_duration: 124.20689655172414
}
{
  _id: 'John Ford',
  total_movies: 29,
  avg_duration: 109.34482758620689
}
}
```

Dado que hay más de 10.000 directores, modifique la consulta para mostrar los 10 primeros con mayor cantidad de películas y su duración promedio.

```
db.movies.aggregate([
  { $unwind: "$directors" },
  { $group: {
    _id: "$directors",
    total_movies: { $sum: 1 },
    avg_duration: { $avg: "$runtime" }
  }
},
{ $sort: { total_movies: -1 } },
{ $limit: 10 },
{ $project: { _id: 0, director: "$_id", total_movies: 1, avg_duration: 1 } }
])
{
  total_movies: 40,
  avg_duration: 96.075,
  director: 'Woody Allen'
}
{
  total_movies: 32,
  avg_duration: 132.84375,
  director: 'Martin Scorsese'
}
{
  total_movies: 31,
  avg_duration: 116.96774193548387,
  director: 'Takashi Miike'
}
{
  total_movies: 29,
  avg_duration: 133.6551724137931,
  director: 'Steven Spielberg'
}
```

```

{
  total_movies: 29,
  avg_duration: 109.34482758620689,
  director: 'John Ford'
}
{
  total_movies: 29,
  avg_duration: 124.20689655172414,
  director: 'Sidney Lumet'
}
{
  total_movies: 27,
  avg_duration: 119.5925925925926,
  director: 'Robert Altman'
}
{
  total_movies: 27,
  avg_duration: 109.88461538461539,
  director: 'Michael Apted'
}
{
  total_movies: 27,
  avg_duration: 130.77777777777777,
  director: 'Clint Eastwood'
}
{
  total_movies: 27,
  avg_duration: 125.55555555555556,
  director: 'Spike Lee'
}

```

5. Distribución de películas por clasificación MPAA

Encuentra la distribución de películas basadas en su clasificación MPAA.

```
> db.movies.aggregate([
  { $group: { _id: "$rated", count: { $sum: 1 } } },
  { $sort: { count: -1 } }
])
< {
  _id: null,
  count: 9894
}
{
  _id: 'R',
  count: 5537
}
{
  _id: 'PG-13',
  count: 2321
}
{
  _id: 'PG',
  count: 1852
}
{
  _id: 'APPROVED',
  count: 709
}
{
  _id: 'G',
  count: 477
}
{
  _id: 'PASSED',
  count: 181
}
{
  _id: 'TV-14',
  count: 89
}
}
```

```
{
  _id: 'TV-PG',
  count: 76
}
{
  _id: 'TV-MA',
  count: 60
}
{
  _id: 'TV-G',
  count: 59
}
{
  _id: 'GP',
  count: 44
}
{
  _id: 'M',
  count: 37
}
{
  _id: 'Approved',
  count: 5
}
{
  _id: 'A0',
  count: 3
}
{
  _id: 'TV-Y7',
  count: 3
}
{
  _id: 'OPEN',
  count: 1
}
```

```
{
  _id: 'Not Rated',
  count: 1
}
```

6. Top 3 países productores de películas

Encuentra los 3 países que producen el mayor número de películas.

```
> db.movies.aggregate([
  { $unwind: "$countries" },
  { $group: { _id: "$countries", count: { $sum: 1 } } },
  { $sort: { count: -1 } },
  { $limit: 3 }
])
< {
  _id: 'USA',
  count: 10921
}
{
  _id: 'UK',
  count: 2652
}
{
  _id: 'France',
  count: 2647
}
```

7. Número promedio de miembros del reparto para películas después del 2000

Para películas lanzadas después del 2000, calcula el número promedio de miembros del reparto.

```
> db.movies.aggregate([
  {
    $match:{
      year: {$gt: 2000},
      cast: {$exists:true, $ne: null},
    },
  },
  {$unwind: "$cast"},
  {
    $group: {
      _id: "$title",
      reparto: {$sum: 1},
    },
  },
  {
    $group:{
      _id: null,
      promedio_miembros_reparto: {$avg: "$reparto"}
    },
  },
  {
    $project: {
      _id: 0,
      promedio_miembros_reparto: {$round: ["$promedio_miembros_reparto", 0]},
    },
  },
])
< {
  promedio_miembros_reparto: 4
}
```

8. Calificación promedio de IMDb y número de comentarios

Determina la calificación promedio de IMDb para las películas y su respectivo número de comentarios.

```
> db.movies.aggregate([
  { $lookup: {
    from: "comments",
    localField: "_id",
    foreignField: "movie_id",
    as: "movie_comments"
  }
},
{ $project: {
  title: 1,
  imdb_rating: "$imdb.rating",
  num_comments: { $size: "$movie_comments" }
}
}]
{
  _id: ObjectId('573a1390f29313caabcd42e8'),
  title: 'The Great Train Robbery',
  imdb_rating: 7.4,
  num_comments: 0
}
{
  _id: ObjectId('573a1390f29313caabcd446f'),
  title: 'A Corner in Wheat',
  imdb_rating: 6.6,
  num_comments: 1
}
{
  _id: ObjectId('573a1390f29313caabcd4803'),
  title: 'Winsor McCay, the Famous Cartoonist of the N.Y. Herald and His Moving Comics',
  imdb_rating: 7.3,
  num_comments: 0
}
```

9. Número total de películas comentadas por usuario

Encuentra el número total de películas en las que cada usuario ha comentado.

```
> db.movies.aggregate([
  {
    $lookup: {
      from: "comments",
      localField: "_id",
      foreignField: "movie_id",
      as: "movie_comments"
    }
  },
  { $project: {
    imdb_rating: {
      $toDouble: {
        $cond: {
          if: { $or: [ { $eq: ["$imdb.rating", ""] }, { $eq: ["$imdb.rating", null] } ] },
          then: 0,
          else: "$imdb.rating"
        }
      }
    },
    num_comments: {
      $size: {
        $ifNull: [ "$movie_comments", [] ]
      }
    }
  }
},
{ $group: {
  _id: null,
  avg_rating: { $avg: "$imdb_rating" },
  avg_comments: { $avg: "$num_comments" }
}
}
])
```

❌ • **MongoServerError(MaxTimeMSExpired):** PlanExecutor error during aggregation :: caused by :: operation exceeded time limit

Excede el tiempo de operación por eso voy a modificar la consulta para tomar las 1000 primeras películas solamente.

```

> db.movies.aggregate([
  {
    $lookup: {
      from: "comments",
      localField: "_id",
      foreignField: "movie_id",
      as: "movie_comments"
    }
  },
  { $project: {
    imdb_rating: {
      $toDouble: {
        $cond: {
          if: { $or: [ { $eq: ["$imdb.rating", ""] }, { $eq: ["$imdb.rating", null] } ] },
          then: 0,
          else: "$imdb.rating"
        }
      }
    },
    num_comments: {
      $size: {
        $ifNull: [ "$movie_comments", [] ]
      }
    }
  }
},
{ $limit: 1000 },
{ $group: {
  _id: null,
  avg_rating: { $avg: "$imdb_rating" },
  avg_comments: { $avg: "$num_comments" }
}
}
])

```

```

< {
  _id: null,
  avg_rating: 7.213,
  avg_comments: 0.77
}

```

10. Películas del género "Western" con comentarios previos a 2020
 Encuentra películas del género "Western" y sus respectivos comentarios antes del "2020-01-01".
 Hice primero una consulta de cuantas películas son, las que cumplen estas condiciones.

```
> db.movies.aggregate([
  { $match: { genres: "Western" } },
  { $lookup: {
    from: "comments",
    localField: "_id",
    foreignField: "movie_id",
    as: "movie_comments"
  }
},
{ $unwind: "$movie_comments" },
{ $match: { "movie_comments.date": { $lt: new ISODate("2020-01-01") } } },
{ $count: "total_western_movies" }
])
< {
  total_western_movies: 96
}
```

```

> db.movies.aggregate([
  { $match: { genres: "Western" } },
  { $lookup: {
    from: "comments",
    localField: "_id",
    foreignField: "movie_id",
    as: "movie_comments"
  }
},
{ $unwind: "$movie_comments" },
{ $match: { "movie_comments.date": { $lt: new ISODate("2020-01-01") } } },
{ $project: { title: 1, "movie_comments.date": 1 } }
])
< {
  _id: ObjectId('573a1391f29313caabcd95f6'),
  title: 'Cimarron',
  movie_comments: {
    date: 2001-04-02T04:43:49.000Z
  }
}
{
  _id: ObjectId('573a1392f29313caabcbda654'),
  title: 'Viva Villa!',
  movie_comments: {
    date: 1973-02-23T15:15:58.000Z
  }
}
{
  _id: ObjectId('573a1393f29313caabcbdd852'),
  title: 'My Darling Clementine',
  movie_comments: {
    date: 1976-09-01T12:57:09.000Z
  }
}
}

```

```
{
  _id: ObjectId('573a1393f29313caababda18'),
  title: 'Bowery Buckaroos',
  movie_comments: {
    date: 1977-03-02T18:10:37.000Z
  }
}
{
  _id: ObjectId('573a1393f29313caababda3'),
  title: 'Fort Apache',
  movie_comments: {
    date: 2005-04-29T12:12:52.000Z
  }
}
{
  _id: ObjectId('573a1394f29313caabcdf2f6'),
  title: 'Carne de horca',
  movie_comments: {
    date: 1992-01-09T12:43:46.000Z
  }
}
{
  _id: ObjectId('573a1394f29313caabcdfd4c'),
  title: 'The Kentuckian',
  movie_comments: {
    date: 2002-04-08T11:51:47.000Z
  }
}
{
  _id: ObjectId('573a1394f29313caabcdfd4c'),
  title: 'The Kentuckian',
  movie_comments: {
    date: 1994-08-18T07:47:24.000Z
  }
}
```

```
{
  _id: ObjectId('573a1394f29313caabcdfdf9'),
  title: 'Oklahoma!',
  movie_comments: {
    date: 1985-12-22T10:28:23.000Z
  }
}
{
  _id: ObjectId('573a1394f29313caabce02ac'),
  title: 'The Rainmaker',
  movie_comments: {
    date: 2015-12-20T05:01:46.000Z
  }
}
{
  _id: ObjectId('573a1394f29313caabce0302'),
  title: 'The Searchers',
  movie_comments: {
    date: 1995-05-29T06:11:50.000Z
  }
}
{
  _id: ObjectId('573a1394f29313caabce031a'),
  title: 'Seven Men from Now',
  movie_comments: {
    date: 1976-10-22T12:46:41.000Z
  }
}
{
  _id: ObjectId('573a1394f29313caabce0993'),
  title: 'The Bravados',
  movie_comments: {
    date: 1990-09-15T04:53:21.000Z
  }
}
```



```
{
  _id: ObjectId('573a1394f29313caabce0c56'),
  title: 'Ride a Crooked Trail',
  movie_comments: {
    date: 1987-06-02T08:37:33.000Z
  }
}
{
  _id: ObjectId('573a1394f29313caabce0c56'),
  title: 'Ride a Crooked Trail',
  movie_comments: {
    date: 2009-10-15T12:11:21.000Z
  }
}
{
  _id: ObjectId('573a1394f29313caabce0fb4'),
  title: 'Last Train from Gun Hill',
  movie_comments: {
    date: 1995-12-06T23:14:35.000Z
  }
}
{
  _id: ObjectId('573a1395f29313caabce1261'),
  title: 'Cimarron',
  movie_comments: {
    date: 1986-12-17T14:14:59.000Z
  }
}
{
  _id: ObjectId('573a1395f29313caabce1811'),
  title: 'The Misfits',
  movie_comments: {
    date: 1971-04-04T14:32:28.000Z
  }
}
```

```
{
  _id: ObjectId('573a1395f29313caabce18f8'),
  title: 'The Second Time Around',
  movie_comments: {
    date: 1988-07-29T10:18:45.000Z
  }
}
{
  _id: ObjectId('573a1395f29313caabce1bfc'),
  title: 'The Man Who Shot Liberty Valance',
  movie_comments: {
    date: 2003-09-25T13:41:24.000Z
  }
}
```

Muestro solo la primer respuesta ya que son 96!