

BASES DE DATOS NOSQL

Carlos Matías Sáez

Máster big data, data science & inteligencia artificial

2023-2024

Universidad Complutense de Madrid



ÍNDICE

1. Introducción
2. Parámetros de la colección de datos
3. Consultas
4. Conclusiones



1. Introducción

En el siguiente trabajo se ha realizado un estudio sobre una colección de datos. En el estudio se ha hecho un análisis de una colección de datos a través de las consultas que nos permite hacer MongoDB junto a su posterior conclusión.

Se ha escogido una colección de datos sobre exoplanetas, la cual se ha obtenido a través del portal “Kaggle”:

<https://www.kaggle.com/datasets/adityamishraml/nasaexoplanets>

Un exoplaneta es un planeta que se encuentra fuera de nuestro sistema solar, orbitando alrededor de una estrella distinta a nuestro Sol. Estos planetas son parte de sistemas estelares diferentes y se descubren mediante observaciones astronómicas, generalmente utilizando telescopios y técnicas de detección que permiten detectar la influencia gravitacional del planeta en su estrella anfitriona o, en algunos casos, observando la disminución de la luz de la estrella cuando un exoplaneta pasa frente a ella (método de tránsito).

La motivación para haber escogido esta colección de datos se debe a que el autor ha estudiado física y es una forma de juntar la rama de los datos con esta.



2. Parámetros de la colección de datos

Un exoplaneta por definición es un planeta que orbita una estrella diferente al Sol y que, por lo tanto, no pertenece al sistema solar. En astrofísica, por lo general, se estudia y se tienen en cuenta una ingente cantidad de parámetros. En este trabajo se ha escogido una colección de datos que contiene parámetros primarios de cada exoplaneta, los cuales son suficientes para caracterizarlos.

En la colección de datos se han incluido todos los exoplanetas descubiertos hasta día de hoy, cuyos parámetros son los siguientes:

- **name:** Nombre del planeta usado por la NASA
- **distance:** Distancia del planeta respecto a la Tierra. Unidades: [ly] (años luz)
- **stellar_magnitude:** Brillo del planeta, cuanto más brillante es el planeta, menor es el número que se le asigna al planeta.
- **planet_type:** Tipo de planeta, estos tipos se derivan de los planetas de nuestro sistema solar.
- **discovery_year:** Año en el que fue descubierto el planeta.
- **mass_multiplier:** Multiplicador de masa del planeta con mass_wrt planet.
- **mass_wrt:** Masa del planeta en comparación con la masa de los planetas de nuestro sistema solar.
- **radius_multiplier:** Multiplicador de radio del planeta con radius_wrt planeta.
- **radius_wrt:** Radio del planeta en comparación con el radio de los planetas de nuestro sistema solar.
- **orbital_radius:** Radio orbital de los planetas que orbitan alrededor de su sol. Unidades: [AU] (unidades astronómicas).
- **orbital_period:** Tiempo en años que tardan esos planetas en completar 1 órbita de su estrella.
- **eccentricity:** Excentricidad. Básicamente indica qué tan circular es la trayectoria orbital, la excentricidad cercana a 0 significa que la trayectoria orbital es casi circular.
- **detection_method:** Método utilizado por la NASA para encontrar ese exoplaneta.



3. Consultas

Una vez importada la colección de datos "*Cleaned_5250*", se realizan una serie de consultas para analizarla.

1. Se quiere saber la cantidad de exoplanetas descubiertos hasta día de hoy.

Consulta:

```
db.Exoplanets.find().count()
```

Resultado:

5250

2. El principal encargado de descubrir exoplanetas entre el año 2009, cuando fue lanzado y el año 2013, donde presentó un fallo y perdió su capacidad de enfocar, fue el telescopio espacial Kepler. Por ello, se realiza la consulta para saber cuantos exoplanetas descubrió consultando los exoplanetas descubiertos entre esos años:

Consulta:

```
db.Exoplanets.count( {discovery_year: { $gte: 2009, $lte: 2013 } } )
```

Resultado:

593

3. Para afinar más la búsqueda y no incluir descubrimientos por otros telescopios, se han incluido los métodos empleados por el telescopio Kepler que son por tránsito y por velocidad radial:

Consulta:

```
db.Exoplanets.count({$and: [
  { discovery_year: { $gte: 2009, $lte: 2013 } },
  { detection_method: { $in: ["Radial Velocity", "Transit"] } }
]})
```



Resultado:

543

4. En el futuro se realizará la primera exploración espacial de un exoplaneta. Las características de este primer exoplaneta serán que se encuentre cerca de la Tierra y que sea del tipo “Terrestrial” (Terrestre) o “Super Earth” (Supertierra). Para ello, busquemos los 11 primeros exoplanetas con estas características:

Consulta:

```
db.Exoplanets.find(
  { $and:
    [ { distance: { $ne: null } },
      { planet_type: { $in: ["Terrestrial", "Super Earth"] } }
    ]
  }).sort({ distance: 1 }).limit(10)
```

Resultado:

Exoplanets		0.030 s	11 Docs			
	_id	name	distance	stellar_magnitude	planet_type	discovery_year
1	6532707bf866d60d4bb	Proxima Centauri b	4.0	11,01.0	Super Earth	2016
2	6532707af866d60d4bb	GJ 887 b	11.0	7,56.0	Super Earth	2020
3	6532707bf866d60d4bb	Ross 128 b	11.0	11,12.0	Super Earth	2017
4	6532707af866d60d4bb	GJ 1061 d	12.0	12,7.0	Super Earth	2020
5	6532707bf866d60d4bb	Tau Ceti e	12.0	3,4961.0	Super Earth	2017
6	6532707bf866d60d4bb	Tau Ceti f	12.0	3,4961.0	Super Earth	2017
7	6532707af866d60d4bb	GJ 1061 b	12.0	12,7.0	Super Earth	2020
8	6532707bf866d60d4bb	Tau Ceti h	12.0	3,4961.0	Super Earth	2017
9	6532707bf866d60d4bb	Tau Ceti g	12.0	3,4961.0	Super Earth	2017
10	6532707af866d60d4bb	GJ 15 A b	12.0	8,09.0	Super Earth	2014
11	6532707af866d60d4bb	GJ 1061 c	12.0	12,7.0	Super Earth	2020

5. La colección de datos no contiene de forma directa la masa del exoplaneta, sino que hay que calcularlo a través de la columna “mass_wrt” y “mass_multiplier”. “mass_wrt” nos muestra el planeta por el que hay que multiplicar “mass_multiplier”, que puede ser Earth o Jupiter. Por tanto, se creará un campo “Mass Earth” que contenga la masa de la Tierra cuando en “mass_wrt” aparezca “Earth” y otro campo “Mass Jupiter” que contenga la masa de Júpiter cuando en “mass_wrt” aparezca “Jupiter”. La masa de la Tierra y de Júpiter son $5,9722 \times 10^{24}$ kg y $1,8981 \times 10^{27}$ kg respectivamente.



Consulta para la Tierra:

```
db.Exoplanets.updateMany(
  { mass_wrt: "Earth" }, // Filtra documentos con "mass_wrt"
  igual a "Earth"
  {
    $set: { "Mass Earth": 5.9722e24 } // Agrega el valor
    constante de la masa de la Tierra
  }
)
```

Consulta para Júpiter:

```
db.Exoplanets.updateMany(
  { mass_wrt: "Jupiter" }, // Filtra documentos con "mass_wrt"
  igual a "Jupiter"
  {
    $set: { "Mass Jupiter": 1.8981e27 } // Agrega el valor
    constante de la masa de Jupiter
  }
)
```

Resultado:

mass_multiplier	mass_wrt	Mass Earth	Mass Jupiter
1,09.0	Earth	5.972.200.000.000.00	
1,14.0	Earth	5.972.200.000.000.00	
0,7.0	Earth	5.972.200.000.000.00	
6,3.0	Jupiter		1.898.100.000.000.00
0,709.0	Jupiter		1.898.100.000.000.00
4,4.0	Jupiter		1.898.100.000.000.00
1,19.0	Jupiter		1.898.100.000.000.00
1,42.0	Jupiter		1.898.100.000.000.00
7,29.0	Jupiter		1.898.100.000.000.00
1,37.0	Jupiter		1.898.100.000.000.00

6. Una vez se tienen creados los campos, se pueden calcular las masas de los planetas creando un nuevo campo *"Mass exoplanet"* multiplicando *"Mass Earth"* o *"Mass Jupiter"* por *"mass_multiplier"*.

Consulta:

```
db.Exoplanets.updateMany(
  { $or: [
    { "Mass Earth": { $exists: true, $ne: null } },
    { "Mass Jupiter": { $exists: true, $ne: null } }
  ] },
  [ { $set: {
    "Mass exoplanet": {
      $cond: [
        { $ifNull: ["$Mass Earth", false] },
        { $multiply: ["$mass_multiplier", "$Mass Earth"] },
        { $multiply: ["$mass_multiplier", "$Mass Jupiter"] }
      ]
    }
  } } ]
)
```

Resultado:

mass_multiplier	mass_wrt	Mass Earth	Mass Jupiter	Mass exoplanet
1,09.0	Earth	5.972.200.000.000.00		6.509.698.000.000.001.000.000.000 (6.5Y)
1,14.0	Earth	5.972.200.000.000.00		6.808.308.000.000.000.000.000.000 (6.8Y)
0,7.0	Earth	5.972.200.000.000.00		4.180.540.000.000.000.000.000.000 (4.2Y)
6,3.0	Jupiter		1.898.100.000.000.00	11.958.029.999.999.999.000.000.000.000 (11.9Y)
0,709.0	Jupiter		1.898.100.000.000.00	1.345.752.900.000.000.000.000.000.000 (1.3Y)
4,4.0	Jupiter		1.898.100.000.000.00	8.351.640.000.000.001.000.000.000.000 (8.3Y)
1,19.0	Jupiter		1.898.100.000.000.00	2.258.738.999.999.999.800.000.000.000 (2.2Y)
1,42.0	Jupiter		1.898.100.000.000.00	2.695.302.000.000.000.000.000.000.000 (2.6Y)
7,29.0	Jupiter		1.898.100.000.000.00	13.837.149.000.000.000.000.000.000.000 (13.8Y)



7. Al igual que con la masa del exoplaneta, para calcular el radio del exoplaneta se hace lo mismo, pero con los campos *"radius_wrt"* y *"radius_multiplier"*. Se creará un campo *"Radius Earth"* y *"Radius Jupiter"*. El radio de la Tierra y de Júpiter es 1.5×10^8 km y 778 412 026 km respectivamente.

Consulta:

```
db.Exoplanets.updateMany(
  { radius_wrt: "Earth" }, // Filtra documentos con "radius_wrt"
                             igual a "Earth"
  {
    $set: { "Radius Earth": 1.5e8 } // Agrega el valor constante
    del radio de la Tierra
  }
)

db.Exoplanets.updateMany(
  { radius_wrt: "Jupiter" }, // Filtra documentos con
                              "radius_wrt" igual a "Jupiter"
  {
    $set: { "Radius Jupiter": 778412026 } // Agrega el valor
    constante del radio de Jupiter
  }
)
```

Resultado:

radius_multiplier ▾	radius_wrt ▾	Radius Jupiter ▾	Radius Earth ▾
1,03.0	Earth		150.000.000 (0.15G)
1,05.0	Earth		150.000.000 (0.15G)
0,913.0	Earth		150.000.000 (0.15G)
1,14.0	Jupiter	778.412.026 (0.78G)	
1,373.0	Jupiter	778.412.026 (0.78G)	
2,07.0	Jupiter	778.412.026 (0.78G)	



8. Una vez tenemos creados ambos campos, podemos calcular los radios de los planetas creando un nuevo campo *"Radius exoplanet"* multiplicando *"Radius Earth"* o *"Radius Jupiter"* por *"Radius_multiplier"*.

Consulta:

```
db.Exoplanets.updateMany(
  { $or: [
    { "Radius Earth": { $exists: true, $ne: null } },
    { "Radius Jupiter": { $exists: true, $ne: null } }
  ] },
  [ { $set: {
    "Radius exoplanet": {
      $cond: [
        { $ifNull: ["$Radius Earth", false] },
        { $multiply: ["$radius_multiplier", "$Radius Earth"] },
        { $multiply: ["$radius_multiplier", "$Radius Jupiter"] }
      ]
    }
  } } ]
)
```

Resultado:

radius_wrt	radius_multiplier	Radius Earth	Radius Jupiter	Radius exoplanet
Earth	1,03.0	150.000.000 (0.15G)		154.500.000 (0.15G)
Earth	1,05.0	150.000.000 (0.15G)		157.500.000 (0.16G)
Earth	0,913.0	150.000.000 (0.15G)		136.950.000 (0.14G)
Jupiter	1,14.0		778.412.026 (0.78G)	887.389.709,64 (0.89G)
Jupiter	1,373.0		778.412.026 (0.78G)	1.068.759.711,698 (1.07G)
Jupiter	2,07.0		778.412.026 (0.78G)	1.611.312.893,82 (1.61G)



9. Ahora que ya tenemos las masas y los radios de los exoplanetas, con motivo de descubrir un exoplaneta con características similares a la Tierra, se filtran aquellos exoplanetas que tengan las siguientes características:

Tipo de planeta "*planet_type*": Terrestre, supertierra.

Masa del exoplaneta "*Mass exoplanet*": [4.7e+24 , 6.2e+24] [kg]

Radio de la órbita "*Radius exoplanet*": [95000000 , 200000000] [km]

Excentricidad "*Eccentricity*": [0.009 , 0.022]. Excentricidad de la tierra: 0.016

Consulta:

```
db.Exoplanets.find({
  $and: [
    { "planet_type": { $in: ["Terrestrial", "Supertierra"] } },
    { "Mass exoplanet": { $gte: 4.7e24, $lte: 6.2e24 } },
    { "Radius exoplanet": { $gte: 95000000, $lte: 200000000 } },
    { "eccentricity": { $gte: 0.009, $lte: 0.022 } }
  ]
})
```

Resultado:

No records found

10. Se quiere saber el año en el que se descubrió la mayoría de los exoplanetas.

Consulta:

```
db.Exoplanets.aggregate([
  {$group: {
    _id: "$discovery_year",
    count: { $sum: 1 }
  },
  {$sort: { count: -1 }
  },
])
```



```

    {$limit: 1
  }
])

```

Resultado:

	_id	count
1	2016	1517 (1.5K)

11. Se quiere obtener el promedio de la masa de todos los exoplanetas "Terrestrial" o "Super Earth":

Consulta:

```

db.Exoplanets.aggregate([
  {$match: {planet_type: { $in: ["Terrestrial", "Super Earth"] } }},
  {$group: {
    _id: null,
    avgMass: { $avg: "$Mass exoplanet" }
  }}
])

```

Resultado:

	_id	avgMass
1	null	31.773.503.820.950.840.000.000.000

*Notar que las unidades de "avgMass" son kg.



12. Se quiere proyectar el nombre, la distancia y el tipo de planeta de los exoplanetas tipo "Gas Giant" descubiertos antes de 2005:

Consulta:

```
db.Exoplanets.aggregate([
  { $match: {
    $and: [
      { planet_type: "Gas Giant" },
      { discovery_year: { $lt: 2005 } }
    ] },
  { $project: {
    name: 1,
    distance: 1,
    planet_type: 1,
    _id: 0
  } } ] )
```

Resultado:

	name	distance	planet_type
1	14 Herculis b	58.0	Gas Giant
2	16 Cygni B b	69.0	Gas Giant
3	2MASS J12073346-3932539 b	210.0	Gas Giant
4	47 Ursae Majoris b	45.0	Gas Giant
5	47 Ursae Majoris c	45.0	Gas Giant
6	51 Pegasi b	50.0	Gas Giant
7	55 Cancri b	41.0	Gas Giant
8	55 Cancri c	41.0	Gas Giant



4. Conclusiones

El primer planeta fue descubierto en 1992, hace relativamente poco tiempo. Por tanto, se puede decir que se han descubierto una amplia cantidad de exoplanetas hasta hoy.

A muchos de los exoplanetas descubiertos le faltan algunas variables todavía por saber, por eso a la hora de trabajar por ejemplo con la distancia a la Tierra se han excluido a aquellos exoplanetas sin valor. Esto es algo a tener en cuenta, ya que puede haber exoplanetas más cercanos que los registrados de los cuales todavía no sabemos su distancia a la tierra.

El primer candidato como exoplaneta a explorar por primera vez sería por ahora "*Proxima Centauri b*" ya que es el más cercano a la Tierra y es del tipo "*Super Earth*". Este tipo de planetas es utilizado para hacer referencia a un planeta terrestre extrasolar que posee entre una y diez veces la masa de la Tierra y entre 1,25 a 2 radios de la Tierra. Los otros tipos de planetas como "*Gas Giant*" o "*Neptune like*" no se han incluido ya que al ser gaseosos no sería posible el aterrizaje de una sonda espacial.

En la consulta 9 se han tomado rangos del valor de los parámetros similares a los de la Tierra. Esto es correcto para encontrar con los parámetros que disponemos, exoplanetas similares a la Tierra, pero realmente a la hora de estudiar estos casos hay una gran variedad de parámetros que no se han tenido en cuenta como el campo magnético o la atmósfera. Por tanto, los resultados que se obtienen son meramente útiles a forma de práctica, no en la realidad.

Después de aplicar un rango notorio en los parámetros de los exoplanetas y no encontrar ningún planeta con esas características, se puede concluir que todavía hace falta que se investiguen y descubran muchos más exoplanetas ya que no se cuenta con el suficiente registro de estos como para que alguno cumpla características similares a las de la Tierra.

