



WORLD
UNIVERSITY
RANKINGS

Uso de MongoDB en un caso de investigacion

BEDU

#data-analysis-21-12

Módulo 1: Introducción a las bases de datos

Diego Oswaldo Gonzalez Macias

Agenda

- Objetivo
- Contexto
- Problema
- Solución
- Conclusiones



Objetivo

Demostrar manejo y dominio de MongoDB:

- Estructura de Base de Datos adecuada, uso debido de datos y practicas coherentes
- Documentos adecuados a la situación
- Consultas simples y un poco complejas
- Uso de vistas



Contexto

Investigación y comparación de las mejores universidades en el mundo y como se encuentran las latinoamericanas, especialmente Mexico.

Fuente de los datos: Kaggle



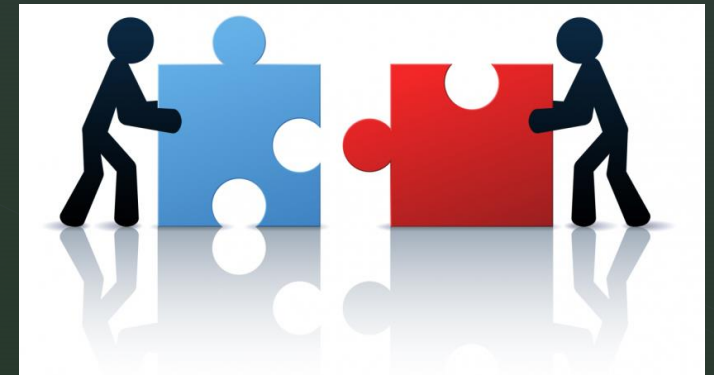
Problema

- Las universidades mexicanas más prestigiosas del país tienen un gran nivel academico, pero ¿será esto cierto o será acaso que estas instituciones se han quedado atrás en comparación con el resto del mundo?
- Se debe identificar donde se encuentra Mexico y a las mejores universidades para buscar enviar ciudadanos prometedores que puedan ayudar al progreso del país y sus instituciones.



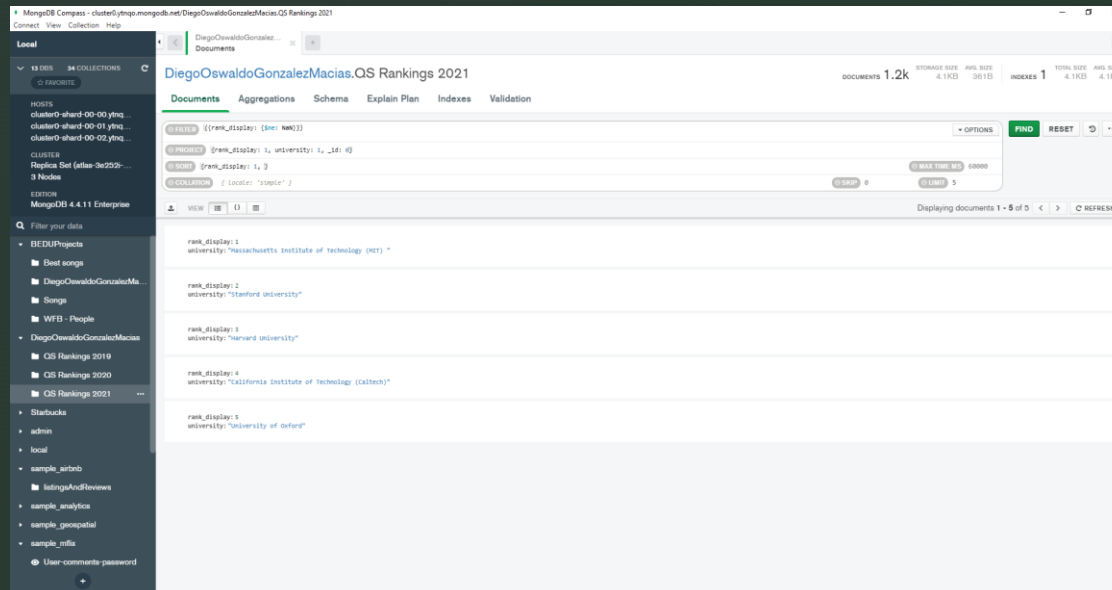
■ Solución

- A. Identificar a las mejores universidades del mundo y sus regiones
- B. Conocer donde se encuentra Latinoamérica y México en el ranking
- C. Observar el impacto de la pandemia en los rankings como posible reflejo de la adaptabilidad de las escuelas internacionales y mexicanas.



A- Identificar a las mejores universidades del mundo y sus regiones

Top 5 Mundial



```
{
  filter: {
    rank_display: {
      $ne: NaN
    }
  },
  project: {
    rank_display: 1,
    university: 1,
    _id: 0
  },
  sort: {
    rank_display: 1
  },
  limit: 5
}
```

Top 30 Norteamerica

rank_display	university	region
1	Massachusetts Institute of Technology (MIT)	North America
2	Stanford University	North America
3	Stanford University	North America
4	Massachusetts Institute of Technology (MIT)	North America
5	University of Chicago	North America
12	Stanford University	North America
16	University of Massachusetts	North America

```
{
  filter: {
    $and: [
      {
        rank_display: {
          $lte: 30
        }
      },
      {
        region: 'North America'
      }
    ]
  },
  project: {
    university: 1,
    rank_display: 1,
    region: 1,
    _id: 0
  },
  limit: 30
}
```


Top 30 Europa

The screenshot shows the MongoDB Compass interface. The left sidebar displays the database structure, including collections like 'best_songs', 'songs', 'wfb_people', and 'qs_rankings_2021'. The main panel shows the 'Documents' tab for the 'DiegoOswaldoGonzalezMacias.QS Rankings 2021' collection. A query is entered in the 'FIND' field: `{ $and: [{ rank_display: { $lte: 30 }, region: 'Europe' }] }`. The results table shows the top 30 universities in Europe, sorted by rank_display. The first few entries are:

rank_display	university	region
1	University of Oxford	Europe
2	ETH Zurich - Swiss Federal Institute of Technology	Europe
3	University of Cambridge	Europe
4	Imperial College London	Europe
5	UCL	Europe
6	University of Edinburgh	Europe

```
{
  filter: {
    $and: [
      {
        rank_display: {
          $lte: 30
        }
      },
      {
        region: 'Europe'
      }
    ]
  },
  project: {
    university: 1,
    rank_display: 1,
    region: 1,
    _id: 0
  },
  limit: 30
}
```

B - Conocer donde se encuentra Latinoamérica y México en el ranking

Promedio Top 10 Latinoamericanas

The screenshot shows the MongoDB Compass interface with an aggregation pipeline applied to the 'OS Rankings 2021' collection. The pipeline consists of three stages: \$limit, \$group, and \$addFields.

```
1. $limit: 10
2. $group: {
  _id: null,
  universities: {
    $sum: 1
  },
  scores: {
    $sum: '$score'
  }
}
3. $addFields: {
  calificacion_promedio_top10: {
    $divide: [
      '$scores',
      '$universities'
    ]
  }
}
```

The output of the \$limit stage shows three sample documents for universities from Argentina, Mexico, and Brazil. The output of the \$group stage shows a single document with an array of universities and their scores. The output of the \$addFields stage shows the calculated average score for the top 10 universities.

```
[$match: {
  region: 'Latin America'
}], {$limit: 10}, {$group: {
  _id: null,
  universities: {
    $sum: 1
  },
  scores: {
    $sum: '$score'
  }
}}, {$addFields: {
  calificacion_promedio_top10: {
    $divide: [
      '$scores',
      '$universities'
    ]
  }
}}
```

Promedio Top 10 mexicanas

The screenshot shows the MongoDB Compass interface with an aggregation pipeline for the 'OS Rankings 2021' collection. The pipeline consists of three stages:

- \$limit: 10**: Limits the documents to 10.
- \$group**: Groups documents by country. The output shows three groups for Mexico, each with a score of 49.9.
- \$addFields**: Adds a new field 'calificacion_promedio_top10' to the output, which is the average of the top 10 scores for each country.

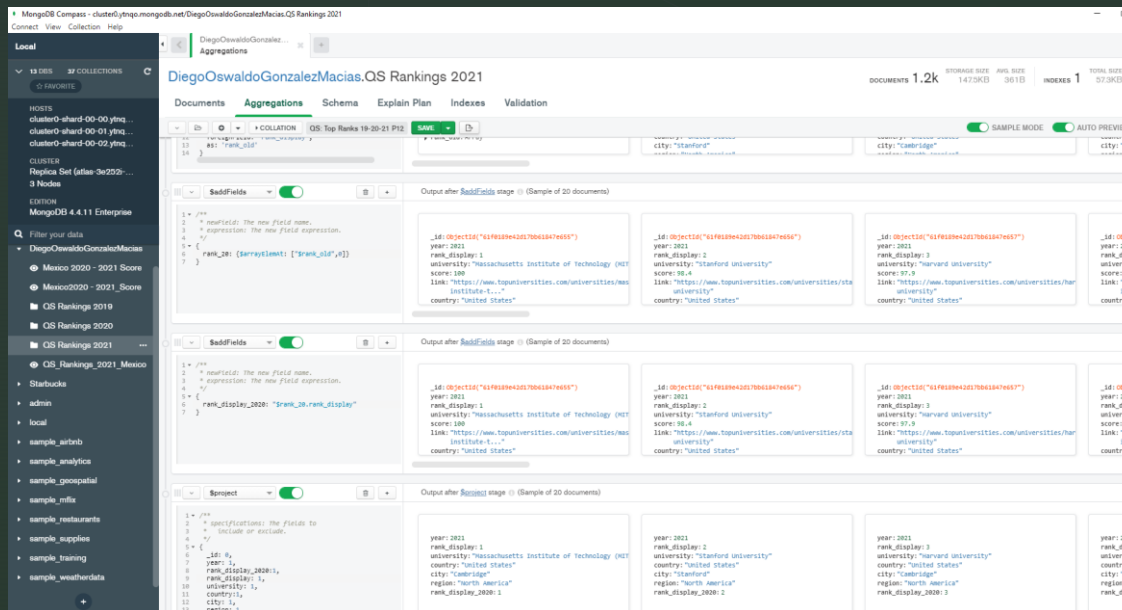
```

[{$match: {
  country: 'Mexico'
}}, {$limit: 10}, {$group: {
  _id: null,
  universities: {
    $sum: 1
  },
  scores: {
    $sum: '$score'
  }
}}, {$addFields: {
  calificacion_promedio_top10: {
    $divide: [
      '$scores',
      '$universities'
    ]
  }
}}]

```

C - Observar el impacto de la pandemia en los rankings

Ranking Top 10 (2020 – 2021)



The screenshot shows the MongoDB Compass interface with an aggregation pipeline for 'DiegoOswaldoGonzalezMacias.QS Rankings 2021'. The pipeline consists of three stages:

- \$setField**: Adds a new field 'rank_display_2020' to the document, setting its value to '\$rank_20.rank_display'.
- \$project**: Projects the fields 'year', 'rank_display', 'rank_display_2020', 'university', 'city', 'country', and 'region'.
- \$lookup**: Performs a lookup operation to find the top 10 universities for each year (2020 and 2021) based on their 'rank_display'.

The output shows a list of universities with their 2020 and 2021 rankings. For example, the first document shows 'Massachusetts Institute of Technology' with a 2020 rank of 1 and a 2021 rank of 1.

```
{ $lookup: {
  from: 'QS Rankings 2020',
  localField: 'rank_display',
  foreignField: 'rank_display',
  as: 'rank_old'
}}, { $addFields: {
  rank_20: {
    $arrayElemAt: [
      '$rank_old',
      0
    ]
  }
}}, { $addFields: {
  rank_display_2020: '$rank_20.rank_display'
}}, { $project: {
  _id: 0,
  year: 1,
  rank_display_2020: 1,
  rank_display: 1,
  university: 1,
  country: 1,
  city: 1,
  region: 1
}}
```

Ranking Top 10 mexicanas (2020 – 2021)

The screenshot shows the MongoDB Compass interface with an aggregation pipeline applied to the 'QS Rankings 2021' collection. The pipeline consists of three stages: \$project, \$sort, and \$match.

```

1 //
2 // $project: the fields to
3 //   include or exclude.
4 //
5 {
6   $project: {
7     "rank_2020": "$rank_2020"
8   }
9 }
10
11 // $sort
12 // $sort: the fields to
13 //   sort by.
14 //
15 {
16   $sort: {
17     "rank_2020": -1
18   }
19 }
20
21 // $match
22 // $match: the query to filter
23 //   documents by.
24 //
25 {
26   $match: {
27     "country": "Mexico"
28   }
29 }

```

The output of the aggregation pipeline shows the top 10 Mexican universities ranked by their 2020 rank. The results are displayed in a table format, showing the university name, its 2020 rank, and its 2021 rank.

University	2020 Rank	2021 Rank
Universidad Nacional Autónoma de México (UNAM)	100	155
Tecnológico de Monterrey	155	171
Universidad Panamericana (UP)	171	185
Universidad del Rosario	185	195
Universidad de los Andes	195	205
Universidad de Chile	205	215
Universidad de Colombia	215	225
Universidad de Venezuela	225	235
Universidad de Argentina	235	245
Universidad de Brasil	245	255

```

[{$match: {
  country: 'Mexico'
}}, {$limit: 10}, {$lookup: {
  from: 'QS Rankings 2020',
  localField: 'rank_display',
  foreignField: 'rank_display',
  as: 'rank_old'
}}, {$addFields: {
  rank_20: {
    $arrayElemAt: [
      '$rank_old',
      0
    ]
  }
}}, {$addFields: {
  rank_2020: '$rank_20.rank_display'
}}, {$project: {
  _id: 0,
  year: 1,
  rank_2020: 1,
  rank_display: 1,
  university: 1,
  country: 1,
  city: 1,
  region: 1
}}]

```


Conclusiones

- Mongo probó ser más fácil de comprender para mí que SQL y eso, junto con el hecho que tiene un desempeño muy bueno, fueron las razones por las que decidí enfocarme en Mongo.
- Las partes mas frustrantes del proyecto fueron: formular las preguntas y generar los JSONs correctos.
- El hecho de que muchas universidades no tengan calificación del año 2020 refleja el impacto que tuvo la pandemia en su decisión de no participar o que QS haya decidido no cambiar las calificaciones.
- El desempeño de las Universidades mexicanas esta muy por detrás, incluso de otras universidades latinoamericanas.
- México debería poner mas interes por mejorar en algunas de las métricas evaluadas para ser mas competitivo.