

# Entregable de Bases de datos NoSQL (MongoDB)

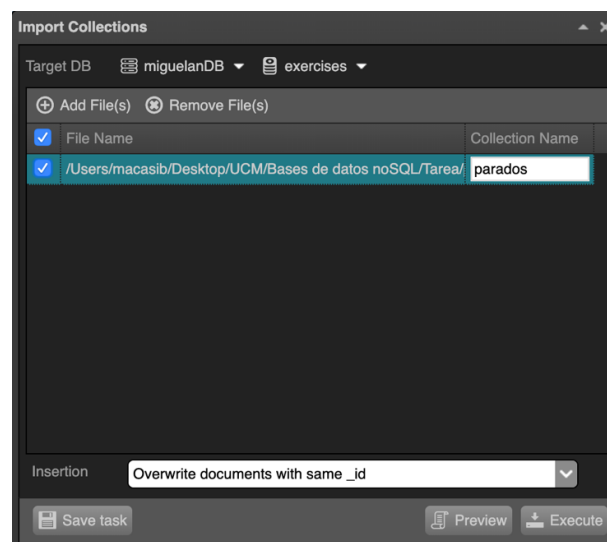
## 1. Introducción

El este trabajo tiene como objetivo el aprendizaje y manejo en bases de datos no relacionales como es MongoDB.

Para ello hemos obtenido un documento en formato JSON desde la pagina [https://datos.gob.es/en/catalogo?res\\_format\\_label=JSON](https://datos.gob.es/en/catalogo?res_format_label=JSON) donde podemos encontrar muchísima información publica para utilizar en este tipo de practicas. En concreto hemos seleccionado un documento publicado por el Instituto Nacional de Estadística donde podemos observar la información de la tasa de desempleo en España separadas en trimestres desde el año 2008, hasta ahora (2020). Este también este clasificado por comunidades autónomas y principales sectores como industria, agricultura, construcción, etc. Vamos a hacer un análisis de como ha evolucionado el desempleo en España estos últimos años teniendo en cuenta la crisis económica de 2008-2015 donde España salió como gran perjudicada. Por otro lado, analizaremos los efectos que ha provocado el COVID-19 en el mercado laboral español.

El documento ha sido modificado para facilitar algunas consultas, borrando algunso campos de meta información, el documento original se puede encontrar en [aquí](#).

Para la inserción de datos he utilizado la herramienta de importación de *NoSQLBooster for Mongo*, que pese a ser mas ineficiente, me pareció una forma mas “friendly” que por terminal. También debemos tener en cuenta que solo hemos importado una colección de 114 documentos, por ello pese a ser mas lento esto ha sido despreciable para este trabajo.



## 2. Documento

Los campos de los que se compone cada documento son:

```
{
  "_id" : "id del objeto (String)",
  "COD" : "código de registro (String)",
  "CCAA" : "Comunidad autónoma (String)",
  "Sector" : "Sector de parados. Total, Agricultura, Construcción,
etc. (String)",
  "T3_Unidad" : "Unidades. En este caso personas (String)",
  "T3_Escala" : "Escala en la que se dan los resultados. Miles de
personas (String)",
  "Data" : [
    {
      "Fecha" : "Fecha de inicio del trimestre.
(timestamp)",
      "T3_TipoDato" : "Definitivo o Provisional (String)",
      "T3_Periodo" : "Trimestre al que pertenece (String)",
      "Anyo" : "Año de los datos (Int)",
      "Valor" : "932.3 (Double)"
    }
  ]
}
```

## 3. CRUD

La principal operación a tener en cuenta en este apartado ha sido el borrado de datos total en la población española, para poder manejar mejor los datos y haciendo consultas con respecto a comunidades autónomas.

- Delete

```
db.parados.deleteMany({$and: [ {"Nombre": /. *Total Nacional.*/} ]})
```

- Update, Create.

Por ahora no ha sido necesario actualizar o añadir ningún campo del documento original, ya que estos eran datos oficiales.

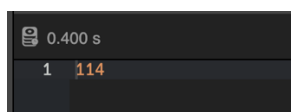
- Read

A continuación, podemos ver las consultas realizadas a la base de datos y proyecciones.

## 4. Queries

- a. En primer lugar, queremos ver el número de documentos, y a continuación listar toda la colección.

```
db.parados.find({}).count()
```



0.400 s	
1	114

```
db.parados.find({})
```

	_id	COD	CCAA	Sector	T3_Unidad	T3_Escala	Data	Nombre
1	5fa531db7765d124bd553c6c	EPA5855	Andalucía	Total CNAE	Personas	Miles	Array[51]	
2	5fa531db7765d124bd553c6d	EPA5854	Andalucía	Agricultura	Personas	Miles	Array[51]	
3	5fa531db7765d124bd553c6e	EPA5853	Andalucía	Industria	Personas	Miles	Array[51]	
4	5fa531db7765d124bd553c6f	EPA5852	Andalucía	Construcción	Personas	Miles	Array[51]	
5	5fa531db7765d124bd553c70	EPA5851	Andalucía	Servicios	Personas	Miles	Array[51]	
6	5fa531db7765d124bd553c71	EPA5850	Andalucía	Parados que busca	Personas	Miles	Array[51]	Parados que buscan
7	5fa531db7765d124bd553c72	EPA5849	Aragón	Total CNAE	Personas	Miles	Array[51]	
8	5fa531db7765d124bd553c73	EPA5848	Aragón	Agricultura	Personas	Miles	Array[51]	
9	5fa531db7765d124bd553c74	EPA5847	Aragón	Industria	Personas	Miles	Array[51]	
10	5fa531db7765d124bd553c75	EPA5846	Aragón	Construcción	Personas	Miles	Array[51]	

- b. Sectores en los que se ha clasificado el desempleo en España. En este ejercicio tomamos como referencia Madrid.

```
db.parados.find({ "CCAA": /. *Madrid.* / }, { "Sector": 1, "T3_Escala": 1, "T3_Unidad": 1, _id: 0 })
```

	Sector	T3_Unidad	T3_Escala
1	Total CNAE	Personas	Miles
2	Agricultura	Personas	Miles
3	Industria	Personas	Miles
4	Construcción	Personas	Miles
5	Servicios	Personas	Miles
6	Parados que buscan primer empleo o han dejado su último empleo hace más de 1 año	Personas	Miles

- c. Trimestre inicial y final en la recolección de datos, Una vez mas aquí tomamos como referencia Madrid

```
db.parados.aggregate([
  { $match: { "CCAA": /. *Madrid.* / } },
  {
    $project:
    {
      _id: 0,
      initDate: { $arrayElemAt: ["$Data.Fecha", 0] },
      lastDate: { $arrayElemAt: ["$Data.Fecha", -1] }
    }
  }
]).limit(1)
```

	initDate	lastDate
1	2020-07-01T00:00:00.000+02:00	2008-01-01T00:00:00.000+01:00

d. Lista de las 5 comunidades que mas desempleados tienen actualmente

```
db.parados.find({
  $and: [
    { "Sector": "Total CNAE" },
  ]
}, { "CCAA": 1, _id: 0 }).sort({ "Data.0.Valor": -1 }).limit(5)
```

	CCAA
1	Andalucía
2	Cataluña
3	Madrid, Comunidad de
4	Comunitat Valenciana
5	Canarias

e. Total de parados el primer trimestre antes y después de la pandemia

```
db.parados.aggregate([
  {
    $match: {
      "Sector": "Total CNAE"
    }
  },
  {
    $project: {
      _id: 0,
      "CCAA": 1,
      "Sector": 1,
      firstTri: { $arrayElemAt: ["$Data.Valor", 0] },
      lastTri: { $arrayElemAt: ["$Data.Valor", 2] },
    }
  },
  { $group: { _id: "Total desempleados España 2020",
    totalFirstTri: { $sum: "$firstTri" }, totalLastTri: { $sum:
    "$lastTri" } }
  })
```

	_id	totalFirstTri	totalLastTri
1	Total desempleados España 2020	3722,8 (3.7K)	3313 (3.3K)

f. Parados en la Región de Murcia clasificados por sector

```
db.parados.aggregate([
  {
    $match: {
      "CCAA": /.Murcia*/
    }
  },
  {
    $project:
    {
      _id: 0,
      "CCAA": 1,
      "Sector": 1,
      Data: { $arrayElemAt: ["$Data", 0] },
    }
  },
  {}
]).sort({ _id: -1 })
```

	CCAA	Sector	Data				
			Fecha	T3_TipoDato	T3_Periodo	Anyo	Valor
1	Murcia, Región de	Total CNAE	2020-07-01T00:00:00.000+02:00	Definitivo	T3	2020 (2.0K)	128,5
2	Murcia, Región de	Agricultura	2020-07-01T00:00:00.000+02:00	Definitivo	T3	2020 (2.0K)	15,2
3	Murcia, Región de	Industria	2020-07-01T00:00:00.000+02:00	Definitivo	T3	2020 (2.0K)	9
4	Murcia, Región de	Construcción	2020-07-01T00:00:00.000+02:00	Definitivo	T3	2020 (2.0K)	2,8
5	Murcia, Región de	Servicios	2020-07-01T00:00:00.000+02:00	Definitivo	T3	2020 (2.0K)	46,9
6	Murcia, Región de	Parados que buscan primer empleo o han dejado st	2020-07-01T00:00:00.000+02:00	Definitivo	T3	2020 (2.0K)	54,6

g. Sectores que más afectados se han visto por la pandemia

– Inicio de la primera ola

```
db.parados.aggregate([
  {
    $project:
    {
      _id: 0,
      "CCAA": 1,
      "Sector": 1,
      Data: { $arrayElemAt: ["$Data", 1] },
    }
  },
  { $group: { _id: "$Sector", totalFirstTri: { $sum: "$Data.Valor" } } }
  {
    $match: {
      _id: { $not: /Total CNAE/ }
    }
  },
  {}
]).sort({ _id: -1 })
```

	_id	totalFirstTri
1	Servicios	1540,4 (1.5K)
2	Parados que buscan primer empleo o han dejado su último empleo hace más de 1 año	1259,6 (1.3K)
3	Industria	231,6
4	Construcción	168
5	Agricultura	168,1

– Inicio de la segunda ola

```
db.parados.aggregate([
  {
    $project:
    {
      _id: 0,
      "CCAA": 1,
      "Sector": 1,
      Data: { $arrayElemAt: ["$Data", 0] },
    }
  },
  { $group: { _id: "$Sector", totalFirstTri: { $sum:
"$Data.Valor" } } }
  {
    $match: {
      _id: { $not: /Total CNAE/ }
    }
  },
  { $sort: { _id: -1 } }
])
```

	_id	totalThidTri
1	Servicios	1561,9 (1.6K)
2	Parados que buscan primer empleo o han dejado su último empleo hace más de 1 año	1597,8 (1.6K)
3	Industria	199,7
4	Construcción	156
5	Agricultura	208

h. Total de parados al inicio de la crisis 2008 (Primer trimestre)

```
db.parados.aggregate([
  {
    $match: {
      "Sector": "Total CNAE"
    }
  },
  {
    $project: {
      _id: 0,
      "CCAA": 1,
      "Sector": 1,
      firstTri: { $arrayElemAt: ["$Data.Valor", -1] },
    }
  },
  { $group: { _id: "Total desempleados España 2008",
    totalFirstTri: { $sum: "$firstTri" } } }
])
```

	_id	totalFirstTri
1	Total desempleados España 2008	2190,6 (2.2K)

i. Total de parados a final de la crisis, en 2015 (Primer trimestre)

```
db.parados.aggregate([
  { $unwind: "$Data" },
  {
    $match: {
      $and: [
        { "Sector": "Total CNAE" }
        { "Data.Anyo": 2015 }
        { "Data.T3_Periodo": "T1" }
      ]
    }
  },
  { $group: { _id: "Total desempleados España 2018",
    totalFirstTri: { $sum: "$Data.Valor" } } }
])
```

	_id	totalFirstTri
1	Total desempleados España 2018	5444,4 (5.4K)

j. Año que mas parados ha tenido España

```
var aux1 = db.parados.aggregate([
  { $unwind: "$Data" },
  {
    $match: {
      $and: [
        { "Sector": "Total CNAE" }
      ]
    }
  }
  { $group: { _id: "$Data.Fecha", totalAnual: { $sum:
"$Data.Valor" } } }
  { $group: { _id: "Máximo parados", maxAnual: { $max:
"$totalAnual" } } }
])

//aux1['maxAnual']
getMaxAnual = function(doc) { return doc.maxAnual; }
var maxAnualAux = aux1.map(getMaxAnual);
//maxAnualAux;

db.parados.aggregate([
  { $unwind: "$Data" },
  {
    $match: {
      $and: [
        { "Sector": "Total CNAE" },
        { "Data.T3_Periodo": "T1" },
      ]
    }
  }
  { $group: { _id: "$Data.Fecha", totalAnual: { $sum:
"$Data.Valor" } } }
  { $match: { "totalAnual": maxAnualAux[0] } } // 6278.1
  { $project: { "totalAnual": 1 } }
])
```

	_id	totalAnual
1	2013-01-01T00:00:00.000+01:00	6278,1 (6.3K)



k. Año que menos parados ha tenido España

```
var aux2 = db.parados.aggregate([
  { $unwind: "$Data" },
  {
    $match: {
      $and: [
        { "Sector": "Total CNAE" }
      ]
    }
  }
  { $group: { _id: "$Data.Fecha", totalAnual: { $sum:
"$Data.Valor" } } }
  { $group: { _id: "Máximo parados", minAnual: { $min:
"$totalAnual" } } }
])

//aux2['minAnual']
getMinAnual = function(doc) { return doc.minAnual; }
var minAnualAux = aux2.map(getMinAnual);
//minAnualAux;

db.parados.aggregate([
  { $unwind: "$Data" },
  {
    $match: {
      $and: [
        { "Sector": "Total CNAE" },
        { "Data.T3_Periodo": "T1" }
      ]
    }
  }
  { $group: { _id: "$Data.Fecha", totalAnual: { $sum:
"$Data.Valor" } } }
  { $match: { "totalAnual": minAnualAux[0] } } //2190.6 }}
  { $project: { "totalAnual": 1 } }
})
```

	_id	totalAnual
1	2008-01-01T00:00:00.000+01:00	2190,6 (2.2K)

## 5. Conclusiones

Tras las consultas que hemos hecho a esta base de datos hemos sacado las siguientes conclusiones:

- El desempleo generado por el coronavirus a ha disminuido, pero teniendo en cuenta esta segunda oleada, puede ser un simple repunte y recaer posteriormente.

- Sectores como la agricultura o servicios se han visto perjudicados, sin embargo, la industria presenta menor tasa de desempleados.
- España es un país desarrollado donde el sector servicio es clave para la economía del país, por tanto, este tipo de empleos son de los mas afectados durante la pandemia y la crisis. En este sector la restauración y el turismo serian piezas clave y donde mas podríamos ver estos efectos si fueran analizados en sectores mas concretos.
- En España las comunidades que mas desempleados reportan son las que mayor población contemplan.
- Además, Murcia presenta una particularidad, y es que el numero de parados en la agricultura es mayor que el de la industria o construcción, esto se debe a ser una zona potencialmente rural donde prima la agricultura en lugar de la industrialización.
- Actualmente en España hay mayor desempleo que al inicio de la crisis de 2008 donde reporto su máximo numero de personas desempleadas en el primer trimestre de 2013.
- Al acabar esta crisis, en 2015, los números no eran notablemente mejores, lo que demuestra que España vivió una recuperación mas lenta que quizás otros países de la Unión Europea.
- También deberíamos de tener en cuenta que la población española a crecido desde 2008 y pese a haber mayor número de personas en el paro, esta tasa por cada mil habitantes podría ser inferior.

A pesar de los datos mostrados en esta colección creo que resultaría de gran ayuda otros documentos tales como la población por comunidad autónoma o una diferenciación por sexo. Esto ayudaría mejor al análisis de estos datos, viendo realmente como actúa el desempleo en España y también podríamos tener la tasa de desempleo por cada 1000 habitantes.