

Intro a Bases de Datos 2020 - BEDU

ALUMNO: Abdiel Céspedes Jaén
EXPERTO: Andrés Ramírez
GRUPO: data-analysis-gdl-20-06



**Breve Interpretación de Bases
de Datos y sus Relaciones**



Objetivos

Objetivo del curso

Al finalizar el módulo serás capaz de utilizar los principales conceptos de las bases de datos mediante la generación de consultas en sistemas gestores de bases de datos relacionales y no relacionales.

Objetivo del proyecto

Resolver las siguientes problemáticas:

- ¿Qué clientes ya han sido satisfechos al haber registrado su estatus como entregado?
- ¿Cuál es el método de pago preferido de los usuarios?
- ¿Cuáles son los pagos que están por encima del promedio de su propio valor?
- ¿Cuál es el valor de pago de cada orden?

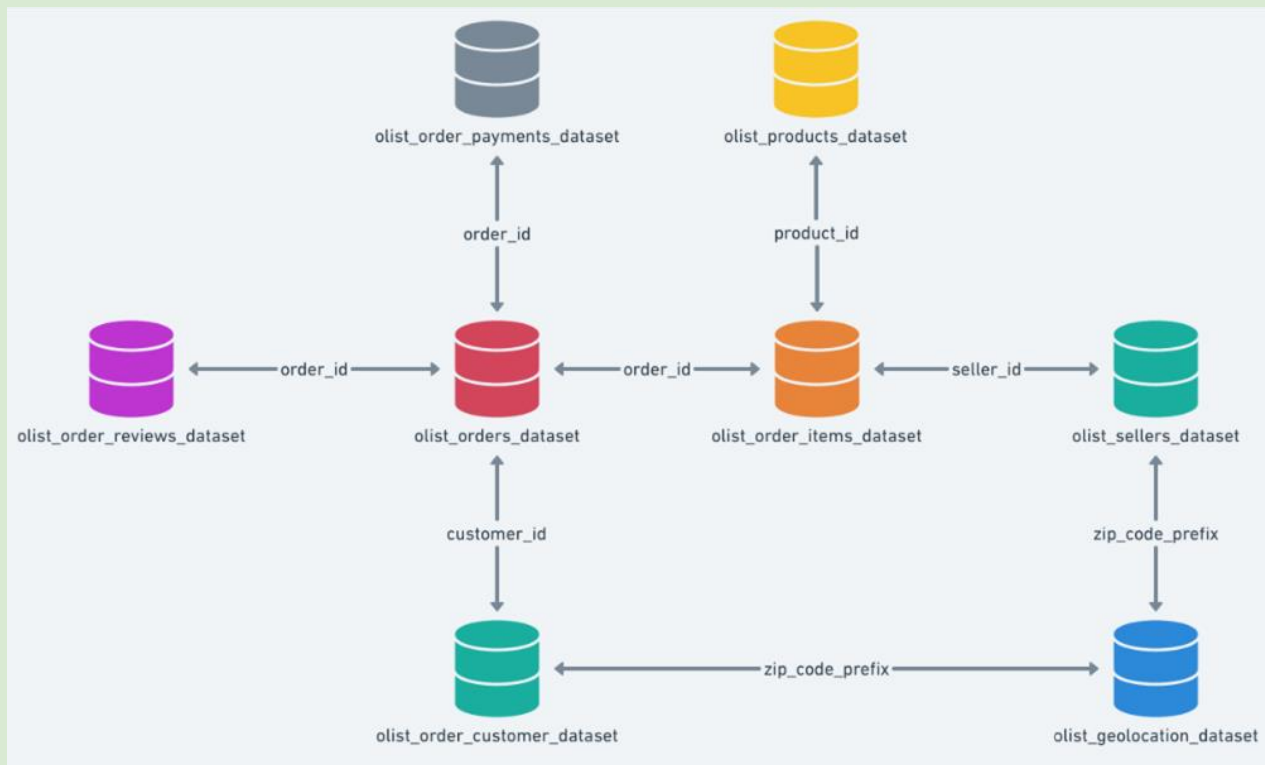
Contexto

La base de datos proviene de una tienda ecommerce brasileña anónima, proveída por www.olist.com. Cuenta con información múltiples dimensiones:

- Estatus de orden.
- Precio.
- Pago.
- Atributos del producto.
- Locaciones del consumidor.
- Reseñas del consumidor.
- Etc.

The Olist logo consists of the word "olist" in a white, lowercase, sans-serif font, centered within a solid blue square.

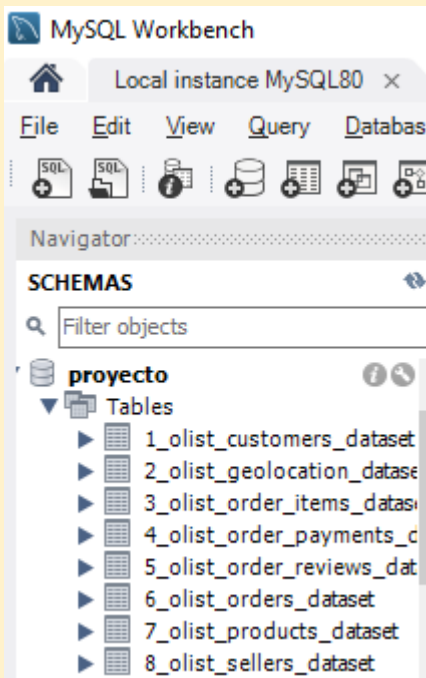
ERR



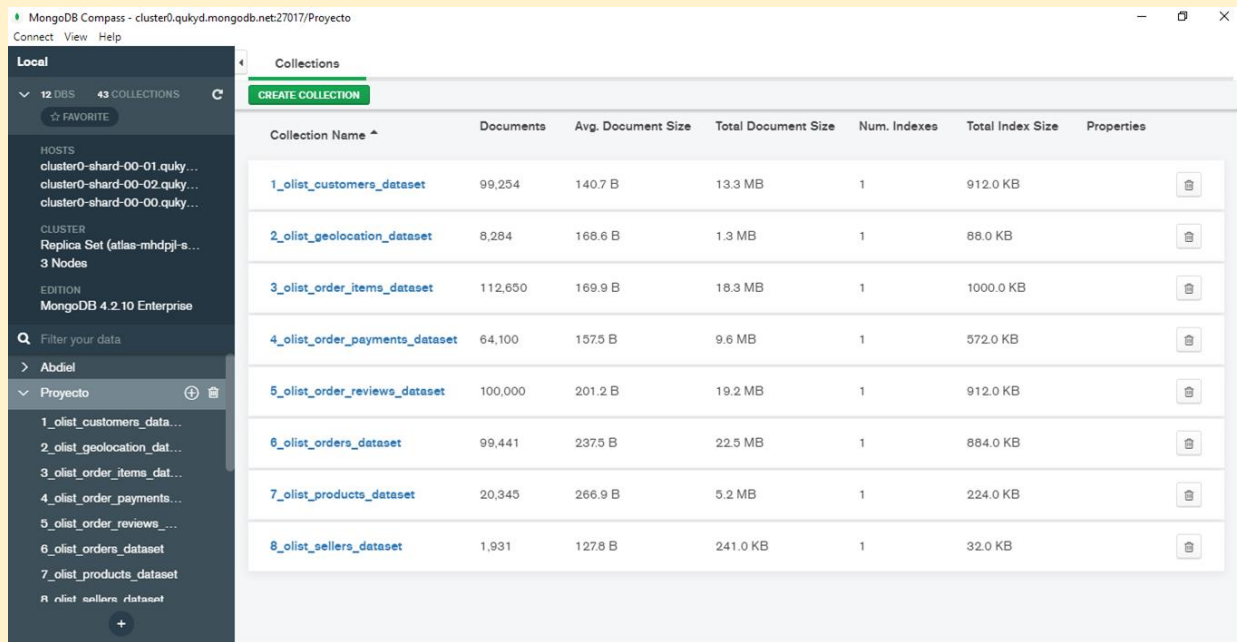


Carga de Bases de Datos

MySQL Workbench



Mongo DB Compass



I. ¿Qué clientes ya han sido satisfechos al haber registrado su estatus de entregado?

	customer_id		
▶	004a2b3e8b026284407c620f4f01f686		
	010583f85dd55d32a6db1068f58ca6f1		
	023d1d56f9187c91230a4e8b50ab6fb3		
	0241bf6728f2c70f948e9a51484d1435		
	02cb822eec8b6309e7ce715ccea3390d		
	02d2c49ef9837fbf18d71940dd7cf97f		
6_olist_orders_dataset 70 ×			
Output			
Action Output ▼			
#	Time	Action	Message
✓ 218	02:31:46	select customer_id, order_status from 6_olist_orders_dataset	500 row(s) returned
✓ 219	02:32:24	select customer_id from 6_olist_orders_dataset where order_status like 'delivered'	487 row(s) returned

FILTER `{ order_status: 'delivered' }`

▼ OPTIONS

FIND

RESET

...

PROJECT `{ _id: 0, customer_id: 1, order_status: 1 }`

SORT

MAXTIMEMS 5000

COLLATION

SKIP 0

LIMIT 0



VIEW



Displaying documents 1 - 20 of 487



REFRESH

```
customer_id: "004a2b3e8b026284407c620f4f01f686"  
order_status: "delivered"
```

```
customer_id: "010583f85dd55d32a6db1068f58ca6f1"  
order_status: "delivered"
```

```
customer_id: "023d1d56f9187c91230a4e8b50ab6fb3"  
order_status: "delivered"
```

```
customer_id: "0241bf6728f2c70f948e9a51484d1435"  
order_status: "delivered"
```

```
customer_id: "02cb822ec8b6309e7ce715ccea3390d"  
order_status: "delivered"
```

II. ¿Cuál es el método de pago preferido por los usuarios?

	payment_type	count(order_id)
▶	credit_card	366
	boleto	97
	voucher	28
	debit_card	8
	not_defined	1

Result 34 ×

Output



Action Output

	#	Time	Action	Message
✓	70	00:24:03	select count(order_id) from 4_olist_order_payments_dataset group by payment_type	5 row(s) returned
✓	71	00:24:19	select payment_type, count(order_id) from 4_olist_order_payments_dataset group by paym...	5 row(s) returned

Proyecto.4_olist_order_payments_dataset

DOCUMENTS 500

TOTAL SIZE 91.1KB

AVG. SIZE 187B

INDEXES 1

TOTAL SIZE 20.0KB

AVG. SIZE 20.0KB

Documents

Aggregations

Schema

Explain Plan

Indexes

Validation



COLLATION

Untitled- Modified

SAVE



SAMPLE MODE



AUTO PREVIEW



```
4  */
5  {
6    _id: 'payment_type',
7    cuenta: {
8      $sum: 1
9    }
10 }
```

_id: boleto
cuenta: 97

_id: credit_card
cuenta: 8



\$sort



Output after \$sort stage (Sample of 5 documents)

```
1  /**
2   * Provide any number of field/order pairs.
3   */
4  {
5    cuenta: -1
6  }
```

_id: "credit_card"
cuenta: 366

_id: "boleto"
cuenta: 97



III. ¿Cuáles son los pagos que están por encima del promedio de su propio valor?

	Cantidad
▶	166

Proyecto.4_olist_order_payments_dataset

DOCUMENTS 500 TOTAL SIZE 91.1KB AVG. SIZE 187B INDEXES 1 TOTAL SIZE 20.0KB AVG. SIZE 20.0KB

Documents

Aggregations

Schema

Explain Plan

Indexes

Validation

COLLATION Untitled - Modified SAVE

SAMPLE MODE AUTO PREVIEW

```
4 {  
5   payment_value: {$gt: 141.8451}  
6 }
```

```
_id: ObjectId("519a1230e0300740483c0923")  
payment_id: 2500  
order_id: "00018f77f2f0320c557190d7a144bdd3"  
payment_sequential: 1  
payment_type: "credit_card"  
payment_installments: 3  
payment_value: 259.83
```

```
_id: ObjectId("519a1230e0300740483c0923")  
payment_id: 12394  
order_id: "000229ec398224ef6ca0657da4fc703e"  
payment_sequential: 1  
payment_type: "credit_card"  
payment_installments: 5  
payment_value: 216.87
```

Count

Output after Count stage (Sample of 1 document)

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

payment_value: 166

IV. ¿Cuál es el valor de pago de cada orden?

	order_id	payment_value
▶	010d14f76a229c295260e4ca4744acc7	44.69
	011375f941b3695c3c3ee58bfe0d3655	144.29
	00bd8cf08fa3998ea6ad6d16c2165822	34.28
	00891ba5de66f55000ee358ceea9b345	149.69
	00ba414462370af2d76195d12e0d5481	39.09
	009838529bb913846ab6670d22865381	176.4

Result 92 ×

Output

Action Output

	#	Time	Action	Message
✓	242	02:46:19	select a.customer_id, b.order_id, c.review_id from 1_olist_customers_dataset a left join 6_ol...	1000 row(s) returned
✓	243	02:47:19	select a.order_id, b.payment_value from 6_olist_orders_dataset a left join 4_olist_order_pay...	258 row(s) returned

Projecto.4_olist_order_payments_dataset

DOCUMENTS 500 TOTAL SIZE 91.1KB AVG. SIZE 187B INDEXES 1 TOTAL SIZE 20.0KB AVG. SIZE 20.0KB

Documents

Aggregations

Schema

Explain Plan

Indexes

Validation

COLLATION Untitled- Modified SAVE SAMPLE MODE AUTO PREVIEW

```
4 * foreignField: The target join field.
5 * as: The name for the results.
6 * pipeline: The pipeline to run on the joined colle
7 * let: Optional variables to use in the pipeline fi
8 */
9 {
10   from: '6_olist_orders_dataset',
11   localField: 'order_id',
12   foreignField: 'order_id',
13   as: 'payment_vale'
14 }
```

```
_id: ObjectId("51501230e0500740483c0928")
payment_id: 85284
order_id: "00010242fe8c5a6d1ba2dd792cb16214"
payment_sequential: 1
payment_type: "credit_card"
payment_installments: 2
payment_value: 72.19
payment_vale: Array
```

```
_id: ObjectId("51501230e0500740483c0928")
payment_id: 2500
order_id: "00018f77f2f0320c557190d7a144bdd3"
payment_sequential: 1
payment_type: "credit_card"
payment_installments: 3
payment_value: 259.83
payment_vale: Array
```

\$project

Output after \$project stage (Sample of 20 documents)

```
1 /**
2  * specifications: The fields to
3  * include or exclude.
4  */
5 {
6   _id: 0,
7   order_id: 1,
8   payment_value: 1
9 }
```

```
order_id: "00010242fe8c5a6d1ba2dd792cb16214"
payment_value: 72.19
```

```
order_id: "00018f77f2f0320c557190d7a144bdd3"
payment_value: 259.83
```



Conclusión

- Gracias a los Sistemas Gestores de Bases de Datos podemos gestionar cantidades abrumadoras de información hasta obtener una interpretación de la realidad.
- Siempre va a haber sesgo y la perspectiva de todas las personas que interactúan con las bases de datos influye en la toma de decisiones.
- Por ello, es necesario apegarse al potencial de estas herramientas de gestión y, en especial, aprenden a abstraer el pensamiento de tal manera que se uno pueda comunicarse directamente con todo un sistema.



Créditos de las imágenes & de bases de datos

Imágenes

- Desconocido. (1 de Junio de 2020). *Para el mercado, el PIB de Brasil caería 6,25%*. Recuperado el 30 de Octubre de 2020, de El Economista: <https://eleconomista.com.ar/2020-06-para-el-mercado-el-pib-de-brasil-caeria-625/>
- Desconocido. (s.f.). *[Imagen]*. Recuperado el 30 de Octubre de 2020, de <https://lh3.googleusercontent.com/leb2NQOJ7aqEyeyrsAtRpRURhAacviecKvWvUh8NIBOk3M8KT6Dr0zM0IXoW6WHYs78>

Base de datos original

- kaggle. (s.f.). *Brazilian E-Commerce Public Dataset by Olist*. Recuperado el 27 de Octubre de 2020, de <https://www.kaggle.com/olistbr/brazilian-ecommerce>

Base de datos utilizada

La base de datos utilizada para la realización de este proyecto se redujo de a 1000 registros/documentos para la primer tabla y a 500 para las demás, esto con el fin de facilitar el proceso de carga de datos a la base de datos local.