

Uso de MySQL y MongoDB en análisis de bases
de datos



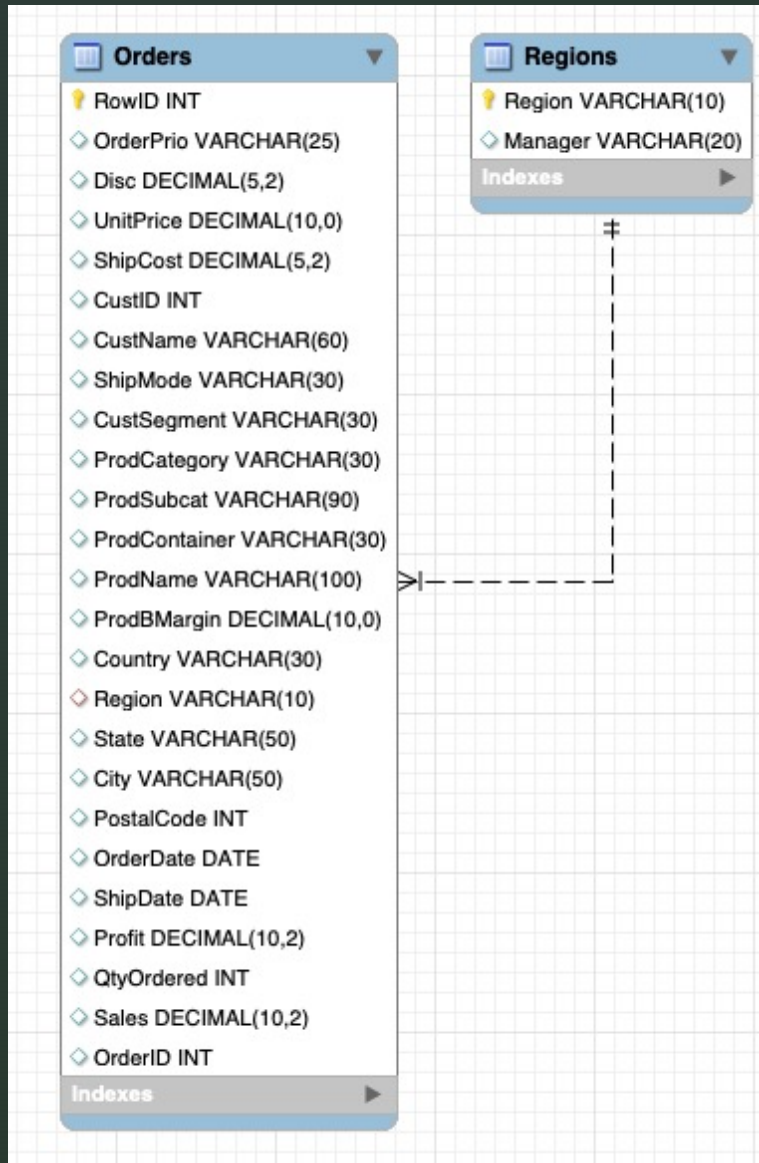
SuperStore US

Daniel Eduardo Lobo Carrillo
data-analysis-21-12

Objetivo

- Comprobar la utilidad y funcionalidad de MySQL y MongoDB para analizar, estructurar y comprender conjuntos grandes de información (datos).
- Realizar diferentes tipos de consultas en ambas herramientas.
- Distinguir en que casos conviene más utilizar bases de datos relacionales o no relacionales.

Diagrama entidad relación



Problema

- Está habiendo muchas órdenes en estado crítico, hay muchas órdenes que están generando pérdida en la región sur y hay muchas ordenes con costo de envío elevado.

Solución

- Identificar el número de órdenes en estado crítico para buscar reducir el número.
- Identificar cuantas órdenes generan pérdidas en la región sur para buscar reducir el número.
- Identificar el menor, mayor y promedio de los costos para buscar un equilibrio.

Identificar el número de órdenes en estado crítico para buscar reducir el número.

- MySQL

```
Select count(*) from orders where  
OrderPrio = "Critical";
```

- MongoDB

```
[{$match: {  
  OrderPrio: 'Critical'  
}}, {$group: {  
  _id: null,  
  OrdenesCriticas: {  
    $sum: 1  
  }  
}}]
```

Identificar el número de órdenes en estado crítico para buscar reducir el número.

```
43 -- 1.- ¿Cuántas ordenes fueron de prioridad crítica?  
44 * Select count(*) from orders where OrderPrio = "Critical";  
45  
46
```

100% 1:44

Result Grid Filter Rows: Search Export:

count(*)

383

MySQL

MongoDB

SuperStoreUS.Orders

DOCUMENTS 2.0k

STORAGE SIZE 360.4KB

AVG. SIZE 575B

INDEXES 1

TOTAL SIZE 32.8KB

Documents Aggregations Schema Explain Plan Indexes Validation

COLLATION Untitled- Modified

SAVE

SAMPLE MODE

AUTO PREVIEW

```
1 /**  
2  * query: The query in MQL.  
3  */  
4 {  
5   OrderPrio: 'Critical'  
6 }
```

```
_id: ObjectId("61f2cd6614fdd2c634815b54")  
RowID: "21776"  
OrderPrio: "Critical"  
Disc: 0.06  
UnitPrice: 9.48  
ShipCost: 7.29  
CustID: "11"  
CustName: "Marcus Dunlap"
```

```
_id: ObjectId("61f2cd6614fdd2c634815b5")  
RowID: "18181"  
OrderPrio: "Critical"  
Disc: 0  
UnitPrice: 4.42  
ShipCost: 4.99  
CustID: "15"  
CustName: "Timothy Reese"
```

```
1 /**  
2  * _id: The id of the group.  
3  * fieldN: The first field name.  
4  */  
5 {  
6   _id: null,  
7   OrdenesCriticas: {  
8     $sum: 1  
9   }  
10 }
```

Output after \$group stage (Sample of 1 document)

```
_id: null  
OrdenesCriticas: 390
```

Identificar cuantas órdenes generan pérdidas en la región sur para buscar reducir el número.

- MySQL

```
Select count(*) from orders  
where Profit < 0 AND Region =  
'South';
```

- MongoDB

```
[{$match: {  
  $and: [  
    {  
      Profit: {  
        $lt: 0  
      }  
    },  
    {  
      Region: 'South'  
    }  
  ]  
}}, {$group: {  
  _id: null,  
  PerdidasSur: {  
    $sum: 1  
  }  
}}]
```


Identificar cuantas órdenes generan pérdidas en la región sur para buscar reducir el número.

```
65 -- 10.- ¿Cuántas órdenes generaron pérdidas de la región Sur?
66 Select count(*) from orders where Profit < 0 AND Region = 'South';
```

100% 16:66

Result Grid Filter Rows: Search Export:

count(*)
230

MySQL

MongoDB

SuperStoreUS.Orders DOCUMENTS 2.0k STORAGE SIZE 360.4KB AVG. SIZE 575B INDEXES 1 TOTAL SIZE 32.8KB AVG. SIZE 32.8KB

Documents Aggregations Schema Explain Plan Indexes Validation

COLLATION Untitled- Modified SAVE SAMPLE MODE AUTO PREVIEW

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

```
1- /**
2-  * query: The query in MQL.
3-  */
4- {
5-   $and: [{Profit:{$lt:0}}, {Region:'South'}]
6- }
```

\$group Output after \$group stage (Sample of 1 document)

```
1- /**
2-  * _id: The id of the group.
3-  * fieldN: The first field name.
4-  */
5- {
6-   _id: null,
7-   PerdidasSur: {
8-     $sum: 1
9-   }
10- }
```

_id: null
PerdidasSur: 233

Identificar el menor, mayor y promedio de los costos para buscar un equilibrio.

- MySQL

```
Select  
MIN(ShipCost),MAX(ShipCost),  
AVG(ShipCost) from orders;
```

- MongoDB

```
[{$group: {  
  _id: null,  
  MenorShipCost: {  
    $min: '$ShipCost'  
  },  
  MayorShipCost: {  
    $max: '$ShipCost'  
  },  
  AvgShipCost: {  
    $avg: '$ShipCost'  
  }  
}]
```

Identificar el menor, mayor y promedio de los costos para buscar un equilibrio.

```
63 -- 9.- ¿Cuál es el menor, mayor y promedio de costo de envío?  
64 • Select MIN(ShipCost),MAX(ShipCost),AVG(ShipCost) from orders;
```

100% 70:60

Result Grid



Filter Rows:



Search

Export:



	MIN(ShipCost)	MAX(ShipCost)	AVG(ShipCost)
--	---------------	---------------	---------------

▶	0.49	164.73	12.787044
---	------	--------	-----------

MySQL

MongoDB

SuperStoreUS.Orders

DOCUMENTS 2.0k STORAGE SIZE 360.4KB AVG. SIZE 575B INDEXES 1 TOTAL SIZE 32.8KB AVG. SIZE 32.8KB

Documents Aggregations Schema Explain Plan Indexes Validation

COLLATION Untitled - Modified SAVE SAMPLE MODE AUTO PREVIEW

Select an operator to construct expressions used in the aggregation pipeline stages. [Learn more](#)

`_id: ObjectId("61f2cd6614fdd2c634815b52")
RowID: "20847"
OrderPrio: "High"
Disc: 0.01
UnitPrice: 2.84
ShipCost: 0.93
CustID: "3"
CustName: "Bonnie Potter"`

`_id: ObjectId("61f2cd6614fdd2c634815b5")
RowID: "20228"
OrderPrio: "Not Specified"
Disc: 0.02
UnitPrice: 500.98
ShipCost: 26
CustID: "5"
CustName: "Ronnie Proctor"`

\$group

Output after \$group stage (Sample of 1 document)

```
1 /**  
2  * _id: The id of the group.  
3  * fieldN: The first field name.  
4  */  
5 {  
6   _id: null,  
7   MenorShipCost: {  
8     $min: "$ShipCost"  
9   },  
10  MayorShipCost: {  
11    $max: "$ShipCost"  
12  },  
13  AvgShipCost: {  
14    $avg: "$ShipCost"  
15  }  
16 }
```

`_id: null
MenorShipCost: 0.49
MayorShipCost: 164.73
AvgShipCost: 12.968150614754098`

Conclusiones

- La base datos se pudo haber desmenuzado más para formar más tablas.
- Las consultas más sencillas en SQL se me hicieron complicadas en MongoDB y viceversa.
- Hay detalles de importación en algunos registros de SQL, en MongoDB es más fácil importar.
- Las consultas tanto en MySQL como Mongo facilitan un análisis más visual y estratégico de conjuntos enormes de datos que difícilmente podrían ser analizados utilizando hojas de cálculo como Excel.