

[h264 @ 0x1c26700] non-existing SPS 0 referenced in buffering period

## Taller MongoDB

URL de acceso a la DB para el taller (estará disponible sólo por 2 semanas hasta que se entregue el parcialito)

mongodb+srv://estudiante:bdd20251c@taller.zxyqk.mongodb.net/?readPreference=secondaryPreferred

URL a una base reducida (siempre disponible)

mongodb+srv://estudiante:bdd20251c@tweets.cbxukni.mongodb.net/

## Consultas Básicas

1. Hallar los tweets del usuario con userid '818839458'

```
db.tweets.find( {user_id: '818839458' } )
```

```
db.tweets.find({user_id: '818839458'})
```

```
db.tweets.find({"user.id_str": '818839458'})
```

2. Hallar aquellos tweets que tengan más de 500000 retweets

```
db.tweets.find({retweet_count: {$gt: 500000}})
```

```
db.tweets.find({retweet_count:{$gt:500000}}, {text:1, 'user.name':1})
```

```
db["tweets"].find({retweet_count: {$gt: 500000}})
```

3. Mostrar la cantidad de retweets de los tweets que se hayan hecho desde Argentina o Brasil.

```
db["tweets"].find({ "user.location": {$in: ["Argentina", "Brasil"]}},  
{retweet_count: 1})
```

```
db.tweets.find({"user.location": {$in: [/Argentina/i, /Brasil/i]}},
```

<code>{retweet_count: 1})</code>
<code>db.tweets.find({"user.location":{\$in:['Argentina','Brasil']}})</code>
<code>db.tweets.find({"place.country": {\$in: [/Argentina/i, /Brasil/i]}}, {retweet_count: 1})</code>
<code>db.tweets.find( {\$or: [{"user.location": /argentina/i}, {"user.location": /brasil/i}]}, {retweet_count: 1})</code>

4. Hallar los usuarios que tengan tweets con 200000 o más retweets y sean en idioma español.

<code>db.tweets.find( {retweet_count: {\$gt: 200000}, lang: 'es'}, {_id: 0, user_id: 1} )</code>
<code>db["tweets"].find({ \$and: [{retweet_count: {\$gt: 200000}}, {"lang": "es"}]}, {user_id: 1})</code>
<code>db.tweets.find({\$and: [{lang: 'es'}, {retweet_count: {\$gte: 200000}}]}, {user_id: 1})</code>
<code>db.tweets.distinct( 'user_id', {retweet_count: {\$gte: 200000}, lang: 'es'}</code> <code>)</code>
<code>db.tweets.find({retweet_count:{\$gte: 200000}, lang:'es'})</code>
<code>db.tweets.find({\$and: [{retweet_count: {\$gte: 200000}}, {lang: 'es'}]}, {"user.name": 1})</code>

5. Mostrar la cantidad de retweets para los tweets que no se hayan hecho en Argentina ni Brasil, pero sí tengan un lugar definido y sean en español.

<code>db["tweets"].find({\$and: [{ "user.location": {\$nin: ["Argentina", "Brasil", null]}}, {"lang": "es"}]}, {_id: 0, retweet_count: 1})</code>
<code>db.tweets.find({\$and: [{"user.location": {\$nin: [/Argentina/, /Brasil/], \$ne: null}}, {lang: 'es'}]}, {retweet_count: 1})</code>
<code>db.tweets.find({   "user.location": {\$nin: [/Argentina/, /Brasil/], \$ne: null},</code>

<pre>    lang: 'es'   }, {retweet_count: 1})</pre>
<pre>db.tweets.find({   "user.location":{\$nin:[/Argentina/,/Brasil/],\$ne:null},   "lang":'es' },{retweet_count:1})</pre>
<pre>db.tweets.find({\$and: [   {"place.country": {\$nin: ["Argentina", "Brasil"]}},   {"place": {\$exists: true}, {lang: 'es'}} ]},{retweet_count: 1})</pre>
<pre>db.tweets.find({"place.country_code":{\$nin:[/AR/,/BR/],\$ne:null},lang:'es' },{retweet_count:1,lang:1,"place.country_code":1})</pre>
<pre>b.tweets.find({\$and: [{"user.location": {\$nin: [/argentina/i, /brasil/i, null]}}, {lang: 'es'}]}), {retweet_count: 1})</pre>
<pre>db.tweets.aggregate([   {\$match: {"user.location": {\$nin: [/Argentina/i, /Brasil/i]}},   {\$match: {"user.location": {\$ne:null}}},   {\$match: {"lang": "es"}},   {\$project: {"retweet_count": 1, "_id": 0}} ])</pre>

6. Mostrar los screen name de aquellos usuarios que tengan “Juan” como parte de su nombre.

<pre>db.tweets.find({"user.name" : /Juan/i}, {"user.screen_name": 1})</pre>
<pre>db.tweets.find({ "user.name": /Juan/i },{ "user.screen_name": 1, _id: 0 })</pre>
<pre>db.tweets.find({"user.name":{\$in:[/Juan/]}},{ "user.screen_name":1,"_id":0} )</pre>
<pre>db.tweets.find(   { "user.name": {\$regex: "Juan", \$options: "i" },   { "user.screen_name": 1, _id: 0 } )</pre>
<pre>db.tweets.aggregate([   {\$match: {"user.name": /Juan/}},</pre>

```
{ $project: { "user.screen_name": 1 } }  
])
```

7. Mostrar de los 10 tweets con más retweets, su usuario y la cantidad de retweets.

```
db.tweets.find({}, {user_id: 1, retweet_count: 1}).sort({retweet_count:  
-1}).limit(10)
```

```
db.tweets.find({}, {user_id: 1, retweet_count: 1}).sort({retweet_count:  
-1}).limit(10)
```

```
db.tweets.find({}, {user_id: 1, retweet_count: 1}).sort({retweet_count:  
-1}).limit(10)
```

```
db.tweets.aggregate([  
  { $sort: { "retweet_count": -1 } },  
  { $limit: 10 },  
  { $project: { "retweet_count": 1, "_id": 0, "user_id": 1 } }  
])
```

## Consultas de Agregación

1. Mostrar de los 10 tweets con más retweets, su usuario y la cantidad de retweets. Ordenar la salida de forma ascendente.

```
db.tweets.aggregate( [  
  { $project: { user_id: 1, retweet_count: 1 } },  
  { $sort: { retweet_count: -1 } },  
  { $limit: 10 },  
  { $sort: { retweet_count: 1 } },  
] )
```

```
db.tweets.aggregate([  
  { $project: { user_id: 1, retweet_count: 1 } },  
  { $sort: { retweet_count: -1 } },  
  { $limit: 10 },  
  { $sort: { retweet_count: 1 } }
```

```
])
```

```
db.tweets.aggregate( [ {$project: {user_id: 1, retweet_count: 1} },  
{$sort: {retweet_count: -1} },{$limit: 10},{$sort: {retweet_count: 1} }])
```

```
db["tweets"].aggregate([  
  {$sort: {retweet_count: -1}},  
  {$limit: 10},  
  {$sort: {retweet_count: 1}},  
  {$project: {retweet_count: 1, "user.name": 1, _id: 0}}])
```

2. Encontrar los 10 hashtags más usados.

```
db.tweets.aggregate( [  
  {$unwind: "$entities.hashtags"},  
  {$group: {_id: "$entities.hashtags.text", veces_usado: {$sum: 1} }  
},  
  {$sort: {veces_usado: -1} },  
  {$limit: 10},  
] )
```

```
db.tweets.aggregate([  
  { $unwind: "$entities.hashtags" },  
  { $group: {  
    _id: "$entities.hashtags.text",  
    count: { $sum: 1 }  
  }},  
  { $sort: { count: -1 } },  
  { $limit: 10 }  
])
```

3. Encontrar los 5 usuarios más mencionados(les hicieron @)

```
db.tweets.aggregate([  
  {$unwind: "entities.user_mentions"},  
  {$group:  
    _id: "$entities.user_mentions.text",  
    count: {$sum:1}
```

```
},
{$sort: {count:-1}},
{$limit: 5}
])
```

```
db.tweets.aggregate( [
  {$unwind: "$entities.user_mentions"},
  {$group: {_id: "$entities.user_mentions.id", menciones: {$sum: 1} } },
  {$sort: {menciones: -1} },
  {$limit: 5},
])
```

```
db.tweets.aggregate([
  {$unwind: "$entities.user_mentions"},
  {$project: {"mencionados": "$entities.user_mentions.id", "_id": 0}},
  {$group: {"_id": "$mencionados", "menciones": {$sum: 1}}},
  {$sort: {"menciones": -1}},
  {$limit: 10},
])
```

4. Hallar la cantidad de retweets promedio para los tweets que se hayan hecho desde Argentina y aquellos que no.

```
db.tweets.aggregate( [
  {$match: {"place.country_code": {$in: ['AR','BR']}}}, {$group: {_id: "$place.country_code",
  retweets: {$sum: "$retweet_count"}}})
```

```
db.tweets.aggregate( [
  {$project: {retweet_count: 1, is_AR: {$eq: ["$place.country_code", "AR"]} } },
  {$group: {_id: "$is_AR", retweet_avg: {$avg: "$retweet_count"}} },
])
```

```
db.tweets.aggregate( {$group: {_id: {$cond: [ {$eq: ["$place.country", "Argentina"]}, "argentina", "other"] }, cantidad_retweets: {$avg: "$retweet_count"} } } )
```

5. Por cada usuario obtener una lista de ids de tweets y el largo de la misma.

```
db.tweets.aggregate( [
```

```
    {$group: {_id: "$user_id", tweets: {$push: "$_id"}} },  
    {$project: {tweets: 1, cant_tweets: {$size: "$tweets"}} },  
  ] )
```

```
db.tweets.aggregate( [  
    {$group: {_id: "$user_id", tweets: {$push: "$_id"}, cantidad: {$count: {}} } },  
  ] )
```

6. Hallar la máxima cantidad de retweets totales que tuvo algún usuario.

```
// O sea, el usuario que recibio la mayor cantidad de retweets totales  
db.tweets.aggregate( [  
    {$group: {_id: "$user_id", retweets: {$sum: "$retweet_count"}} },  
    c,  
    {$limit: 1},  
  ] )
```

```
db.tweets.aggregate([  
  {$group: {_id: "$user.screen_name", "retweets": {$sum: "$retweet_count"}}},  
  {$sort: {"retweets": -1}},  
  {$limit: 1})
```

7. Hallar para cada intervalo de una hora cuántos tweets realizó cada usuario.

```
db.tweets.aggregate( [  
    {$group: {  
      _id: {"user": "$user_id", "hora": {$hour: "$created_at"}},  
      cant: {$sum: 1},  
    } },  
  ] )
```

```
db.tweets.aggregate( [  
    {$group: {  
      _id: {  
        "user": "$user_id",  
        "hora": {$hour: "$created_at"},
```

```
        "dia": {$dayOfMonth:"$created_at"},
        "mes": {$month:"$created_at"},
        "anio": {$year:"$created_at"}
    }, count:{$sum: 1}
  }} ])
```

## *Un tweet de ejemplo:*

```
{
  "_id": "1143676626808557568",
  "created_at": {
    "$date": "2019-06-26T00:25:30.000Z"
  },
  "full_text": "https://t.co/gTtsYlX7cH",
  "truncated": false,
  "display_text_range": [
    0,
    0
  ],
  "entities": {
    "hashtags": [],
    "symbols": [],
    "user_mentions": [],
    "urls": [],
    "media": [
      {
        "id": {
          "$numberLong": "1143676619707621376"
        },
        "id_str": "1143676619707621376",
        "indices": [
```



```

        0,
        23
    ],
    "media_url": "http://pbs.twimg.com/media/D98n03-XoAA1a5a.jpg",
    "media_url_https":
"https://pbs.twimg.com/media/D98n03-XoAA1a5a.jpg",
    "url": "https://t.co/gTtsYlX7cH",
    "display_url": "pic.twitter.com/gTtsYlX7cH",
    "expanded_url":
"https://twitter.com/capturebizarre/status/1143676626808557568/photo/1
",
    "type": "photo",
    "sizes": {
        "thumb": {
            "w": 150,
            "h": 150,
            "resize": "crop"
        },
        "small": {
            "w": 639,
            "h": 452,
            "resize": "fit"
        },
        "medium": {
            "w": 639,
            "h": 452,
            "resize": "fit"
        },
        "large": {
            "w": 639,
            "h": 452,
            "resize": "fit"
        }
    }
}
}
]
},

```

```
"source": "<a href=\"http://twitter.com/download/iphone\"  
rel=\"nofollow\">Twitter for iPhone</a>",  
"in_reply_to_status_id": null,  
"in_reply_to_status_id_str": null,  
"in_reply_to_user_id": null,  
"in_reply_to_user_id_str": null,  
"in_reply_to_screen_name": null,  
"user": {  
  "id_str": "818839458",  
  "name": "CAPTURE BIZARRE",  
  "screen_name": "capturebizarre",  
  "location": "Argentina",  
  "description": "• Nos gustan los memes •  
comercial@capturebizarre.com •",  
  "url": "https://t.co/S1MZ7Uk0Ib",  
  "protected": false,  
  "followers_count": 1426721,  
  "friends_count": 423,  
  "listed_count": 1393,  
  "created_at": "Wed Sep 12 04:43:59 +0000 2012",  
  "favourites_count": 6888,  
  "time_zone": null,  
  "verified": true,  
  "statuses_count": 49159,  
  "lang": null,  
  "default_profile": false  
},  
"geo": null,  
"coordinates": null,  
"place": null,  
"contributors": null,  
"is_quote_status": false,  
"retweet_count": 18404,  
"favorite_count": 53515,  
"possibly_sensitive": false,  
"lang": "und",  
"text": "https://t.co/gTtsYlX7cH",  
"user_id": "818839458",
```

```
"in_user_hashtag_collection": true,  
"hashtag_origin_checked": true,  
"cooccurrence_checked": true  
}
```