

MongoDB

Part 1:

1. Download and Insert this dataset to a MongoDB collection under a database.
**mongoimport --db mydb --collection Electronics --type json --file
'/home/cloudera/Downloads/Electronics_5.json'**
2. Use 'OR' operation to show only the 'summary' and 'overall' columns with overall rating 1 or 3 or 5.

**db.Electronics.find({\$or:[{"overall": 1}, {"overall": 3}, {"overall": 5}],
{ "summary":1,"overall":1}).pretty()**

```
Type "it" for more
> db.Electronics.find({$or:[{"overall": 1}, {"overall": 3}, {"overall": 5}], {"summary":1,"overall":1,"_id":0}).pretty()
{ "overall" : 5, "summary" : "Gotta have GPS!" }
{ "overall" : 1, "summary" : "Very Disappointed" }
{ "overall" : 3, "summary" : "1st impression" }
{ "overall" : 1, "summary" : "Major issues, only excuses for support" }
{ "overall" : 5, "summary" : "HDMI Nook adapter cable" }
{ "overall" : 5, "summary" : "A Perfedct Nook HD+ hook up" }
{ "overall" : 5, "summary" : "This works great but read the details..." }
{ "overall" : 5, "summary" : "great fit" }
{ "overall" : 5, "summary" : "Works Great" }
{ "overall" : 3, "summary" : "It Works" }
{ "overall" : 5, "summary" : "Great replacement for original power cable" }
{
  "overall" : 3,
  "summary" : "This is the oem charger you get with the nook hd+"
}
{ "overall" : 3, "summary" : "$45 for a power cord? Good luck with that." }
{
  "overall" : 5,
  "summary" : "Excelant mount for Tv. Would buy more if I needed them."
}
{ "overall" : 5, "summary" : "Perfect" }
{ "overall" : 5, "summary" : "32 In TV Mount" }
{ "overall" : 5, "summary" : "Cant Beat it!" }
{ "overall" : 5, "summary" : "Just what I'd hoped" }
{ "overall" : 5, "summary" : "Nice Mount" }
{ "overall" : 5, "summary" : "VideoSecu Swing Arm TV Wall Mount" }
Type "it" for more
> db.Electronics.find({$or:[{"overall": 1}, {"overall": 3}, {"overall": 5}], {"summary":1,"overall":1,"_id":0}).pretty().count()
1260008
> █
```

3. Use 'AND' operation to show only the 'reviewerName' and 'overall' columns with overall rating less than 3 (use \$lt) and greater than 1 (use \$gt) and sort the result by the ascending alphabetic order of 'reviewerName'.

**db.Electronics.find({\$and:[{'overall': {\$gt : 1}}, {'overall': {\$lt: 3}}]}, {'overall':1,
'reviewerName': 1,'_id':0}).sort({'reviewerName': 1}).pretty()**

```
db.Electronics.find({'$and':[{'overall': {$gt : 1}}, {'overall': {$lt: 3}}]}, {'overall':1, 'reviewerName': 1, '_id':0}).sort({'reviewerName': 1}).pretty()
{ "overall" : 2 }
{ "overall" : 2 }
{ "overall" : 2 }
{ "overall" : 2 }
{ "overall" : 2 }
{ "overall" : 2 }
{ "overall" : 2 }
{ "overall" : 2 }
{ "overall" : 2 }
{ "overall" : 2 }
{ "overall" : 2 }
{ "overall" : 2 }
{ "overall" : 2 }
{ "overall" : 2 }
{ "overall" : 2 }
{ "overall" : 2 }
{ "overall" : 2 }
{ "overall" : 2 }
{ "overall" : 2 }
Type "it" for more
> db.Electronics.find({'$and':[{'overall': {$gt : 1}}, {'overall': {$lt: 3}}]}, {'overall':1, 'reviewerName': 1, '_id':0}).sort({'reviewerName': 1}).pretty().count()
82139
>
```

Part 2:

1. Find the 'reviewText' that have the word 'awesome'.

```
db.Electronics.find({'reviewText':{'regex':'awesome'}},{'reviewText':1,
'_id':0}).pretty()
```

```

File Edit View Search Terminal Help
cently become available, and I'm glad I got my Nook when I did; I like it so much better than ANY ereader I've looked at. I can drop my nook in my purse, and anytime I'm stuck somewhere and want to read, I've got a whole library right there with me. The Nook is awesome!!!!"
}
{
    "reviewText": "I have been super happy with this reader. In the bright sun sitting outside, no problem. On the many flights I take, no problem. B&N; site and downloading books is a breeze. I've got a Droid X phone, and the accounts sync, so I can quickly pull up the same book on my phone with the Nook App, very cool.I can lend, and have others lend me, some of the books purchased.Our local library in Central Texas, outside Austin, is connected to Overdrive.com. With my library card I can download epub books and check them out, anywhere I've got internet. I don't have to check them back in, they just go away after the check out period. Same with Audiobooks. Nook handles this format, other eReaders do not, so you might want to look into this option and what formats the readers have before making a decision.I like the onscreen keyboard/touch screen, personally. It is bright and easy in areas where dimly lit and the light in your reader case is directed at the reading screen. Seems like no big deal, but after 28 hours on a plane you get tired and this makes a difference.Not a bunch of keys to hit/press with the keyboard exposed. This is a personal opinion, and I can absolutely understand why others may want the keyboard. I put in pagemarks etc on books, but don't type a lot of notes into the ebooks when reading. If you do a lot of this, you may want to test run the onscreen keyboard and see if you like it. I don't, the amount of typing is very limited, so onscreen is perfect and I'm used to it with my phones.Using this over your wifi to surf the internet is something you can do, but it isn't the best. You can in a pinch, but it's easier for me to pull out my smartphone and crank into a website.The number of places to get books is, especially free ones, is awesome with the Nook handling so many formats. B&N; just search 8.00 and titles and you'll get thousands of classics come up that copyright is over and they are free. gutenber.org is puts free classics by the tens of thousands out that you can download onto your Nook for free, google books does this and a few other sites. Tons and Tons of books for free, I haven't read Tom Sawyer in 20+ years and am now, for free. Just an addition to note.Lots of periodicals, newspapers etc, that you can get daily.I take my kids into B&N; and they love to look through the kids books. While there, I can read any book over their wifi for free for up to an hour. I don't use this a lot, but once a month or so when I take the kids, it's pretty cool.Overall, I'm really psyched, this Nook has been awesome."
}
{
    "reviewText": "I cannot believe the seller who sold me this not good, great Nook HD+, that is keeping me very happy. I have been truly blessed, this Nook tablet is the best. I am watching movies on it via Hulu and Netflix and other apps for movies etc. The video and audio are great for me either outside on my porch or indoors.I do not plan to do much more other than playing a game here and there, this tablet is sufficient to my needs. There is no camera, if you need one get another type of tablet. I recommend for those who do not need a super tablet, however, this tablet can do a lot. I just have not tried it all yet. I will again say this, the video playback to me is awesome,it is truly HD. I am very satisfied with my refurbished Nook HD+, I am getting a lot for very little. I recommend buying thNook for a few dollars worth a lot of fun."
}
{
    "reviewText": "Here's what you want to know.The stand is FINE. There are only 5 working buttons on this. This is the power button, a brightness button(s), and volume buttons. There is no other screens, setups, no calibrations, etc. All of that has to be done on your PC connected to this. It is literally a straight monitor. As advertised, not a dead pixel.Resolutions excellent.There is a powerbrick with this item (just an FYI). Haven't had any issues with it.To mount the monitor on the monitor mount, put the backend on first, then put the big base on the 4 feet, screw them in (recommend using a drill on a very low setting to affix it). The DVI connector is a good length.I'd recommend it. Went from a dual monitors at 1920x1080 (24 inch) to this beast. Night and day. I can have two documents up side by side with room to spare (my girl reviews allot of documents, and she was liking this hunking thing too). It is a \"Large\" monitor (comparable to a 30 inch TV).Yes the instructions are in Korean, who cares. The box has the whole spare parts theme on it, don't care either. It is a good monitor, good response time, awesome resolution.I would do this monitor at 2560x1600 vs a 4k monitor for 4 reasons.1) 4k monitors are $1,000+. Yes there are cheaper ones but there cheaper because...2) Most 4k monitors run at 30hz. Basically it means that the monitor can only display 30hz. THIS monitor or displays at 60 hz (which is fantastic). To clarify, essentially hz is pretty much how much fps a monitor can display per second. Now you can have a 4k monitor but if your graphics card is kicking at 1000 fps, your monitor will only support 30 fps of it (its a dumbed down version of it but hz matter, and you want at least 60, if not more)3) 4k monitors may require more than 1 connection (aka 2 dvi port connections) to run well.4) 4k is still reaaaally new. There working the bugs out regarding creating a gaming 4k monitor that isn't over $1000. This is a 30 inch monitor, well under the 700 to 800 price range of other monitors that do the same thing as this one.To the gamers out there, loaded up counter strike, awesomenauts, eve online, all ran at this resolution, full screen, no bars."
}
}
Type "it" for more
> db.Electronics.find({'reviewText':{'$regex':'awesome'}},{ 'reviewText':1, '_id':0}).pretty().count()
38135
>

```

2. Find the 'summary' that have any characters apart form 'alphanumerical characters' and 'space'.

```
db.Electronics.find({'summary' : {'$not' : {'$regex' : '[A-Za-z0-9\s]'}}}).pretty()
```

```

cloudera@quickstart:~
File Edit View Search Terminal Help

"asin" : "B0002AKX6Y",
"reviewerName" : "75mrclean",
"helpful" : [
  0,
  0
],
"reviewText" : "This product worked for about 12 hours then it burned out. I guess you get what you pay for. Not very bright light either.",
"overall" : 1,
"summary" : "?????????",
"unixReviewTime" : 1359072000,
"reviewTime" : "01 25, 2013"
}
{
  "_id" : ObjectId("61748ad12c72347f8d8efc71"),
  "reviewerID" : "A2LZQFKLAAXFFE",
  "asin" : "B0002J1LEI",
  "reviewerName" : "bduphs",
  "helpful" : [
    0,
    0
  ],
  "reviewText" : "Just what the doctor ordered for a great price. Worked as intended and received in record time. Would recommend to all.",
  "overall" : 5,
  "summary" : "!!!",
  "unixReviewTime" : 1345248000,
  "reviewTime" : "08 18, 2012"
}
{
  "_id" : ObjectId("61748ad12c72347f8d8f07dd"),
  "reviewerID" : "ALFHKWTIUBAIS",
  "asin" : "B0002L5R78",
  "reviewerName" : "Briana Villalobos \"SnowWhite\"",
  "helpful" : [
    0,
    0
  ],
  "reviewText" : "Men and their toys. Husband needed this, works great for him and good price as well. So I recommend it. =)",
  "overall" : 5,
  "summary" : "=)",
  "unixReviewTime" : 1365465600,
  "reviewTime" : "04 9, 2013"
}
Type "it" for more
> db.Electronics.find({'summary' : {$not : {$regex:'[A-Za-z0-9\s]'}}}).pretty().count()
333

```

3. Use \$nin operator to show only the 'summary' and 'overall' columns where overall rating not in 2 and 4.
4. **db.Electronics.find({'overall': {\$nin: [2,4]}}, {'summary':1,'overall':1,'_id':0}).pretty()**

```

---
> db.Electronics.find({'overall': {$nin: [2,4]}}, {'summary':1,'overall':1,'_id':0}).pretty()
{ "overall" : 5, "summary" : "Gotta have GPS!" }
{ "overall" : 1, "summary" : "Very Disappointed" }
{ "overall" : 3, "summary" : "1st impression" }
{ "overall" : 1, "summary" : "Major issues, only excuses for support" }
{ "overall" : 5, "summary" : "HDMI Nook adapter cable" }
{ "overall" : 5, "summary" : "A Perfected Nook HD+ hook up" }
{ "overall" : 5, "summary" : "This works great but read the details..." }
{ "overall" : 5, "summary" : "great fit" }
{ "overall" : 5, "summary" : "Works Great" }
{ "overall" : 3, "summary" : "It Works" }
{ "overall" : 5, "summary" : "Great replacement for original power cable" }
{
  "overall" : 3,
  "summary" : "This is the oem charger you get with the nook hd+"
}
{ "overall" : 3, "summary" : "$45 for a power cord? Good luck with that." }
{
  "overall" : 5,
  "summary" : "Excelant mount for Tv. Would buy more if I needed them."
}
{ "overall" : 5, "summary" : "Perfect" }
{ "overall" : 5, "summary" : "32 In TV Mount" }
{ "overall" : 5, "summary" : "Cant Beat it!" }
{ "overall" : 5, "summary" : "Just what I'd hoped" }
{ "overall" : 5, "summary" : "Nice Mount" }
{ "overall" : 5, "summary" : "VideoSecu Swing Arm TV Wall Mount" }
Type "it" for more
> db.Electronics.find({'overall': {$nin: [2,4]}}, {'summary':1,'overall':1,'_id':0}).pretty().count()
1260008
>

```

5. Show only the 'reviewerID' and 'reviewerName' column with the ascending order of 'reviewerID'. Show only first 10 result.

db.Electronics.find({},{"reviewerID":1,"reviewerName":1,'_id':0}).sort({reviewerID:1}).pretty().limit(10)

```
Type "it" for more
> db.Electronics.find({},{"reviewerID":1,"reviewerName":1,'_id':0}).sort({reviewerID:1}).pretty().limit(10)
{ "reviewerID" : "A000715434M800HLCENK9", "reviewerName" : "DP" }
{ "reviewerID" : "A000715434M800HLCENK9", "reviewerName" : "DP" }
{ "reviewerID" : "A000715434M800HLCENK9", "reviewerName" : "DP" }
{ "reviewerID" : "A000715434M800HLCENK9", "reviewerName" : "DP" }
{ "reviewerID" : "A000715434M800HLCENK9", "reviewerName" : "DP" }
{ "reviewerID" : "A00101847G3FJTWYGNQA", "reviewerName" : "Cang Cheng" }
{ "reviewerID" : "A00101847G3FJTWYGNQA", "reviewerName" : "Cang Cheng" }
{ "reviewerID" : "A00101847G3FJTWYGNQA", "reviewerName" : "Cang Cheng" }
{ "reviewerID" : "A00101847G3FJTWYGNQA", "reviewerName" : "Cang Cheng" }
{ "reviewerID" : "A00101847G3FJTWYGNQA", "reviewerName" : "Cang Cheng" }
```

Part 3:

1. Group by 'reviewerName' to show the minimum 'overall' rating they posted. Show only 10 results.

db.Electronics.aggregate([{\$group: {_id: "\$reviewerName", overall: {\$min: "\$overall"}}}, { \$limit : 10 }]).pretty()

```
> db.Electronics.aggregate([{$group: {_id: "$reviewerName", overall: {$min: "$overall"}}}, { $limit : 10 }]).pretty()
{ "_id" : "Timothy Chu", "overall" : 4 }
{ "_id" : "Bayou Reader", "overall" : 4 }
{ "_id" : "HiFiGuy528", "overall" : 2 }
{ "_id" : "A. Hung", "overall" : 3 }
{ "_id" : "Michael D. Leach", "overall" : 4 }
{ "_id" : "J. Easton", "overall" : 1 }
{ "_id" : "rehpotsirhc", "overall" : 2 }
{ "_id" : "eyeinzsky75", "overall" : 1 }
{ "_id" : "prettyinparadise", "overall" : 4 }
{ "_id" : "Jason M Parks", "overall" : 2 }
```

2. Group by 'Helpful' to show the 'total number' of entries found for different 'Helpful' data, sorted by descending order of the 'total number'.

db.Electronics.aggregate([{\$group: { '_id': '\$helpful', 'totalnumber': {\$sum:1}}}, {\$sort: {'totalnumber':-1}}]).pretty()

```

> db.Electronics.aggregate([{$group: { '_id': '$helpful', 'totalnumber': {$sum:1} }}, {$sort: {'totalnumber': -1}}]).pretty()
{ "_id" : [ 0, 0 ], "totalnumber" : 963227 }
{ "_id" : [ 1, 1 ], "totalnumber" : 201597 }
{ "_id" : [ 2, 2 ], "totalnumber" : 68757 }
{ "_id" : [ 0, 1 ], "totalnumber" : 67890 }
{ "_id" : [ 1, 2 ], "totalnumber" : 45798 }
{ "_id" : [ 3, 3 ], "totalnumber" : 33519 }
{ "_id" : [ 2, 3 ], "totalnumber" : 23580 }
{ "_id" : [ 4, 4 ], "totalnumber" : 19214 }
{ "_id" : [ 3, 4 ], "totalnumber" : 16225 }
{ "_id" : [ 0, 2 ], "totalnumber" : 13453 }
{ "_id" : [ 5, 5 ], "totalnumber" : 12106 }
{ "_id" : [ 1, 3 ], "totalnumber" : 9688 }
{ "_id" : [ 4, 5 ], "totalnumber" : 9367 }
{ "_id" : [ 6, 6 ], "totalnumber" : 8456 }
{ "_id" : [ 2, 4 ], "totalnumber" : 7367 }
{ "_id" : [ 5, 6 ], "totalnumber" : 6202 }
{ "_id" : [ 7, 7 ], "totalnumber" : 6050 }
{ "_id" : [ 3, 5 ], "totalnumber" : 5406 }
{ "_id" : [ 0, 3 ], "totalnumber" : 4910 }
{ "_id" : [ 6, 7 ], "totalnumber" : 4485 }
Type "it" for more

```

- Find the count of data where 'unixReviewTime' greater than a certain value (try different values of 'unixReviewTime' from the dataset to see if the result changes. Submit any of the result you got for your input value of 'unixReviewTime').

Value1:

```
db.Electronics.find({"unixReviewTime": {$gt: 1290643200}}).count()
```

```
result = 1405419
```

Value2:

```
db.Electronics.find({"unixReviewTime": {$gt: 1283990400}}).count()
```

```
result = 1429732
```

Value3:

```
db.Electronics.find({"unixReviewTime": {$gt: 1370131200}}).count()
```

```
result = 671219
```

```

> db.Electronics.find({"unixReviewTime": {$gt: 1290643200}}).count()
1405419
> db.Electronics.find({"unixReviewTime": {$gt: 1283990400}}).count()
1429732
> db.Electronics.find({"unixReviewTime": {$gt: 1370131200}}).count()
671219
> █

```