

**Practical No . 2**

**Aim :** Practical of Data collection, Data curation and management for Large-scale Data system (such as MongoDB) .

Theory :

- **Data Collection –**

- **Data collection** is the process of gathering and [measuring information](#) on targeted variables in an established system, which then enables one to answer relevant questions and evaluate outcomes.
- [Data](#) collection is a [research](#) component in all study fields, including [physical](#) and [social sciences](#), [humanities](#), and [business](#).
- The goal for all data collection is to capture quality evidence that allows analysis to lead to the formulation of convincing and credible answers to the questions that have been posed.

- **Data Curation –**

- **Data curation** is the organization and integration of [data](#) collected from various sources.
- It involves annotation, publication and presentation of the data such that the value of the data is maintained over time, and the data remains available for reuse and preservation.
- Data curation includes "all the processes needed for principled and [controlled data](#) creation, maintenance, and [management](#), together with the capacity to add value to data".

- **Data Management –**

- **Data management** is the practice of collecting, keeping, and using data securely, efficiently, and cost-effectively.
- The goal of data management is to help people, organizations, and connected things optimize the use of data within the bounds of policy and regulation so that they can make decisions and take actions that maximize the benefit to the organization.
- A robust data management strategy is becoming more important than ever as organizations increasingly rely on intangible assets to create value.

Data (csv format) :

Serial_code	Mobile_Model	Operating_System	Internal Storage	Ram	Rear Camera Pixel
7575	Samsung M31	Android 10	64 GB	3 GB	20 MP
4234	Apple iPhone 12	iOS 14	512 GB	4GB	24 MP
1639	Moto G5	Android 8	64 GB	4 GB	26 MP
7364	Samsung Galaxy S21	Android 10	128 GB	6 GB	64 MP
8648	One Plus 8	Android 11	64 GB	8 GB	32 MP
1289	Apple iPhone 6	iOS 9	128 GB	2 GB	8 MP
1209	Redmi Note 9	Android 10	128 GB	6 GB	13 MP
6324	RealMe 3	Android 10	128 GB	3 GB	20 MP
8644	Apple iPhone XR	iOS 12	64 GB	2 GB	12 MP
8430	Asus ROG 2	Android 10	128 GB	8 GB	20 MP

**Steps :**

1. First import the csv file data using the following command :

Syntax – **mongoimport -d database\_name(db name we want to keep) -c collection\_name(collection name we want to keep) --type data\_type(file type)--file file\_name(with extension) --headerline**

```

Administrator: C:\Windows\system32\cmd.exe
Microsoft Windows [Version 6.1.7601]
Copyright (c) 2009 Microsoft Corporation. All rights reserved.

C:\Users\Administrator>cd..
C:\Users>cd..
C:\>cd C:\mongodb-win32-x86_64-2012plus-4.2.12\bin
C:\mongodb-win32-x86_64-2012plus-4.2.12\bin>mongoimport -d database2 -c scores --
--type csv --file mobileinfo.csv --headerline
2021-02-22T13:37:19.297+0530 connected to: mongodb://localhost/
2021-02-22T13:37:19.559+0530 10 document(s) imported successfully. 0 document
(s) failed to import.
C:\mongodb-win32-x86_64-2012plus-4.2.12\bin>_

```

2. Run the following Command to view the databases we have : (cmd - **show dbs**)

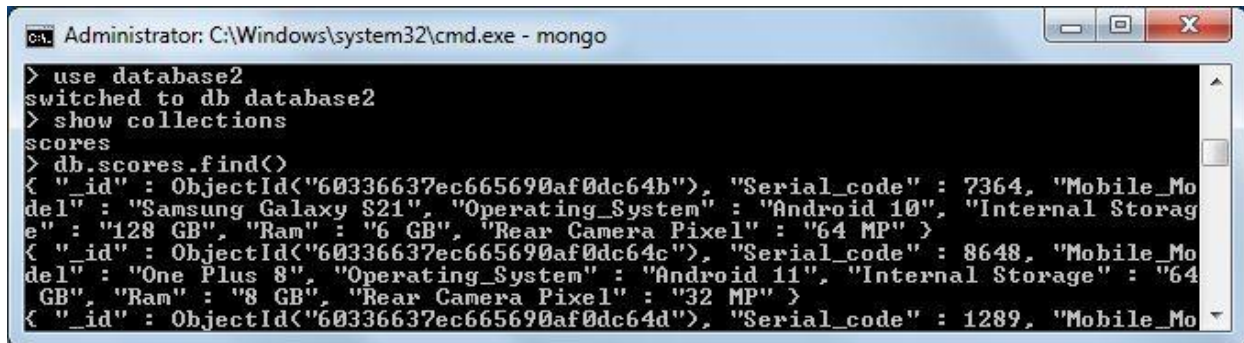
```

MongoDB>
> show dbs
admin          0.000GB
company        0.000GB
config         0.000GB
database2     0.000GB
local          0.000GB
> _

```

3. Now we can use our database :

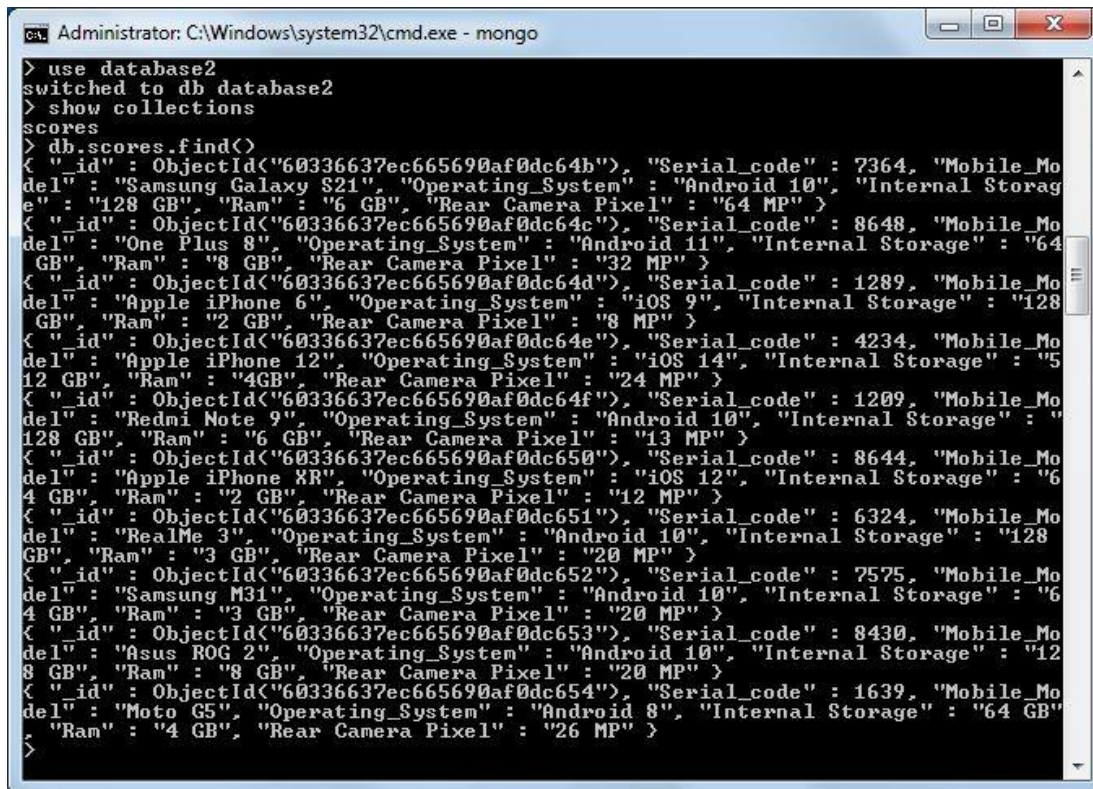
Syntax : **use database\_name**



```
Administrator: C:\Windows\system32\cmd.exe - mongo
> use database2
switched to db database2
> show collections
scores
> db.scores.find()
{ "_id" : ObjectId("60336637ec665690af0dc64b"), "Serial_code" : 7364, "Mobile_Model" : "Samsung Galaxy S21", "Operating_System" : "Android 10", "Internal_Storage" : "128 GB", "Ram" : "6 GB", "Rear_Camera_Pixel" : "64 MP" }
{ "_id" : ObjectId("60336637ec665690af0dc64c"), "Serial_code" : 8648, "Mobile_Model" : "One Plus 8", "Operating_System" : "Android 11", "Internal_Storage" : "64 GB", "Ram" : "8 GB", "Rear_Camera_Pixel" : "32 MP" }
{ "_id" : ObjectId("60336637ec665690af0dc64d"), "Serial_code" : 1289, "Mobile_Model" : "Apple iPhone 6", "Operating_System" : "iOS 9", "Internal_Storage" : "128 GB", "Ram" : "2 GB", "Rear_Camera_Pixel" : "8 MP" }
```

4. Now we have to get the collections we have in our database for that we have to run the following command :

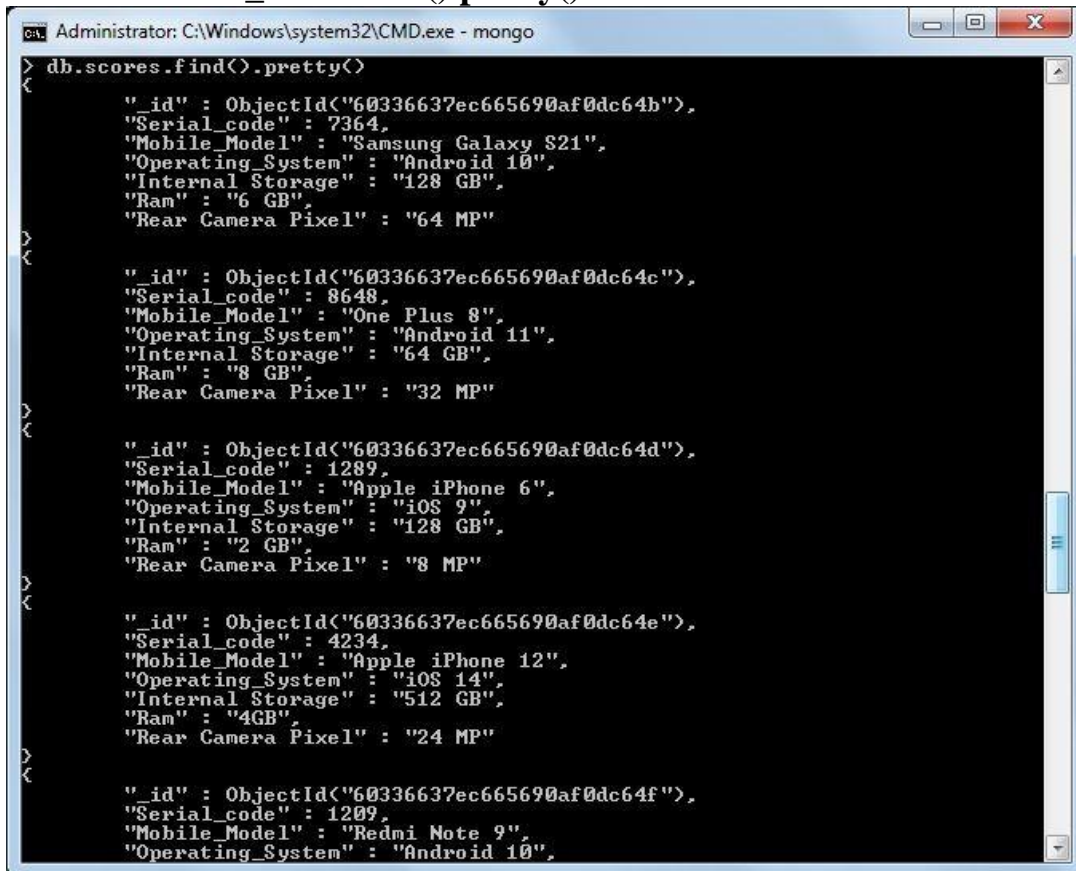
Syntax : **show collections**



```
Administrator: C:\Windows\system32\cmd.exe - mongo
> use database2
switched to db database2
> show collections
scores
> db.scores.find()
{ "_id" : ObjectId("60336637ec665690af0dc64b"), "Serial_code" : 7364, "Mobile_Model" : "Samsung Galaxy S21", "Operating_System" : "Android 10", "Internal_Storage" : "128 GB", "Ram" : "6 GB", "Rear_Camera_Pixel" : "64 MP" }
{ "_id" : ObjectId("60336637ec665690af0dc64c"), "Serial_code" : 8648, "Mobile_Model" : "One Plus 8", "Operating_System" : "Android 11", "Internal_Storage" : "64 GB", "Ram" : "8 GB", "Rear_Camera_Pixel" : "32 MP" }
{ "_id" : ObjectId("60336637ec665690af0dc64d"), "Serial_code" : 1289, "Mobile_Model" : "Apple iPhone 6", "Operating_System" : "iOS 9", "Internal_Storage" : "128 GB", "Ram" : "2 GB", "Rear_Camera_Pixel" : "8 MP" }
{ "_id" : ObjectId("60336637ec665690af0dc64e"), "Serial_code" : 4234, "Mobile_Model" : "Apple iPhone 12", "Operating_System" : "iOS 14", "Internal_Storage" : "512 GB", "Ram" : "4GB", "Rear_Camera_Pixel" : "24 MP" }
{ "_id" : ObjectId("60336637ec665690af0dc64f"), "Serial_code" : 1209, "Mobile_Model" : "Redmi Note 9", "Operating_System" : "Android 10", "Internal_Storage" : "128 GB", "Ram" : "6 GB", "Rear_Camera_Pixel" : "13 MP" }
{ "_id" : ObjectId("60336637ec665690af0dc650"), "Serial_code" : 8644, "Mobile_Model" : "Apple iPhone XR", "Operating_System" : "iOS 12", "Internal_Storage" : "64 GB", "Ram" : "2 GB", "Rear_Camera_Pixel" : "12 MP" }
{ "_id" : ObjectId("60336637ec665690af0dc651"), "Serial_code" : 6324, "Mobile_Model" : "RealMe 3", "Operating_System" : "Android 10", "Internal_Storage" : "128 GB", "Ram" : "3 GB", "Rear_Camera_Pixel" : "20 MP" }
{ "_id" : ObjectId("60336637ec665690af0dc652"), "Serial_code" : 7575, "Mobile_Model" : "Samsung M31", "Operating_System" : "Android 10", "Internal_Storage" : "64 GB", "Ram" : "3 GB", "Rear_Camera_Pixel" : "20 MP" }
{ "_id" : ObjectId("60336637ec665690af0dc653"), "Serial_code" : 8430, "Mobile_Model" : "Asus ROG 2", "Operating_System" : "Android 10", "Internal_Storage" : "128 GB", "Ram" : "8 GB", "Rear_Camera_Pixel" : "20 MP" }
{ "_id" : ObjectId("60336637ec665690af0dc654"), "Serial_code" : 1639, "Mobile_Model" : "Moto G5", "Operating_System" : "Android 8", "Internal_Storage" : "64 GB", "Ram" : "4 GB", "Rear_Camera_Pixel" : "26 MP" }
```

5. For extracting the documents from collections we have the following command :

Syntax - **db.collection\_name.find().pretty()**



```
Administrator: C:\Windows\system32\CMD.exe - mongo
> db.scores.find().pretty()
{
  "_id" : ObjectId("60336637ec665690af0dc64b"),
  "Serial_code" : 7364,
  "Mobile_Model" : "Samsung Galaxy S21",
  "Operating_System" : "Android 10",
  "Internal_Storage" : "128 GB",
  "Ram" : "6 GB",
  "Rear_Camera_Pixel" : "64 MP"
}
{
  "_id" : ObjectId("60336637ec665690af0dc64c"),
  "Serial_code" : 8648,
  "Mobile_Model" : "One Plus 8",
  "Operating_System" : "Android 11",
  "Internal_Storage" : "64 GB",
  "Ram" : "8 GB",
  "Rear_Camera_Pixel" : "32 MP"
}
{
  "_id" : ObjectId("60336637ec665690af0dc64d"),
  "Serial_code" : 1289,
  "Mobile_Model" : "Apple iPhone 6",
  "Operating_System" : "iOS 9",
  "Internal_Storage" : "128 GB",
  "Ram" : "2 GB",
  "Rear_Camera_Pixel" : "8 MP"
}
{
  "_id" : ObjectId("60336637ec665690af0dc64e"),
  "Serial_code" : 4234,
  "Mobile_Model" : "Apple iPhone 12",
  "Operating_System" : "iOS 14",
  "Internal_Storage" : "512 GB",
  "Ram" : "4GB",
  "Rear_Camera_Pixel" : "24 MP"
}
{
  "_id" : ObjectId("60336637ec665690af0dc64f"),
  "Serial_code" : 1209,
  "Mobile_Model" : "Redmi Note 9",
  "Operating_System" : "Android 10",
}
```

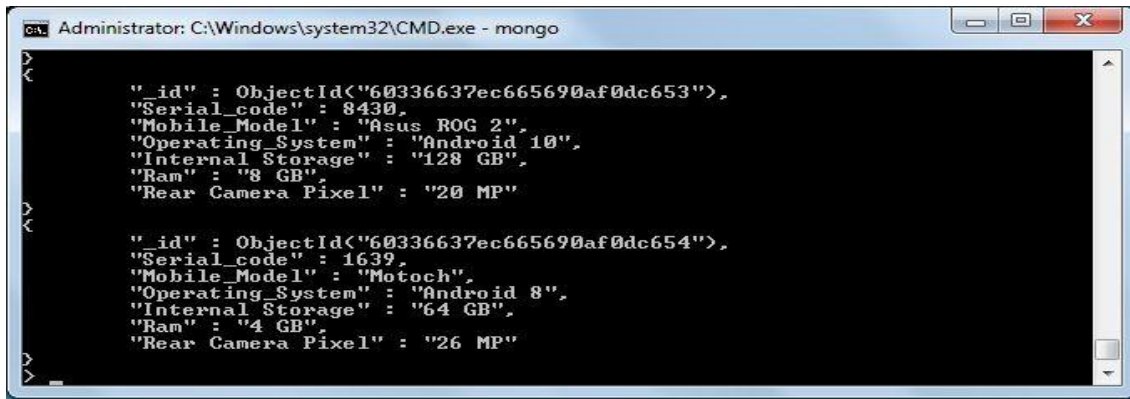
6. Now we can perform basic mongodb commands to manipulate the data in our database. The command we would be using first is Update command .

Syntax – **db.collection\_name.update(selection\_criteria,update\_criteria)**



```
Administrator: C:\Windows\system32\CMD.exe - mongo
...
> db.scores.update(
...   { "Serial_code" : 1639 },
...   { $set: { "Mobile_Model" : "Motoch" } } )
WriteResult<{ "nMatched" : 1, "nUpserted" : 0, "nModified" : 1 }>
> -
```

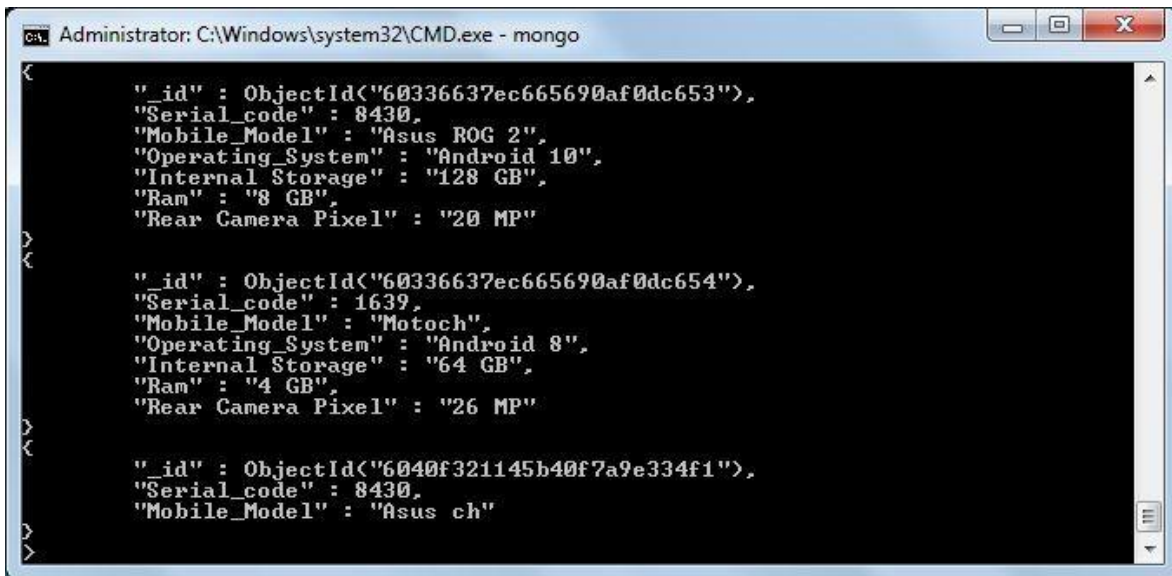




```

Administrator: C:\Windows\system32\CMD.exe - mongo
>
{
  "_id" : ObjectId<"60336637ec665690af0dc653">,
  "Serial_code" : 8430,
  "Mobile_Model" : "Asus ROG 2",
  "Operating_System" : "Android 10",
  "Internal_Storage" : "128 GB",
  "Ram" : "8 GB",
  "Rear_Camera_Pixel" : "20 MP"
}
{
  "_id" : ObjectId<"60336637ec665690af0dc654">,
  "Serial_code" : 1639,
  "Mobile_Model" : "Motoch",
  "Operating_System" : "Android 8",
  "Internal_Storage" : "64 GB",
  "Ram" : "4 GB",
  "Rear_Camera_Pixel" : "26 MP"
}
>

```



```

Administrator: C:\Windows\system32\CMD.exe - mongo
{
  "_id" : ObjectId<"60336637ec665690af0dc653">,
  "Serial_code" : 8430,
  "Mobile_Model" : "Asus ROG 2",
  "Operating_System" : "Android 10",
  "Internal_Storage" : "128 GB",
  "Ram" : "8 GB",
  "Rear_Camera_Pixel" : "20 MP"
}
{
  "_id" : ObjectId<"60336637ec665690af0dc654">,
  "Serial_code" : 1639,
  "Mobile_Model" : "Motoch",
  "Operating_System" : "Android 8",
  "Internal_Storage" : "64 GB",
  "Ram" : "4 GB",
  "Rear_Camera_Pixel" : "26 MP"
}
{
  "_id" : ObjectId<"6040f321145b40f7a9e334f1">,
  "Serial_code" : 8430,
  "Mobile_Model" : "Asus ch"
}
>

```

7.Next command we will be using is remove command :

Syntax : **db.collection\_name.remove(deletion\_criteria)**

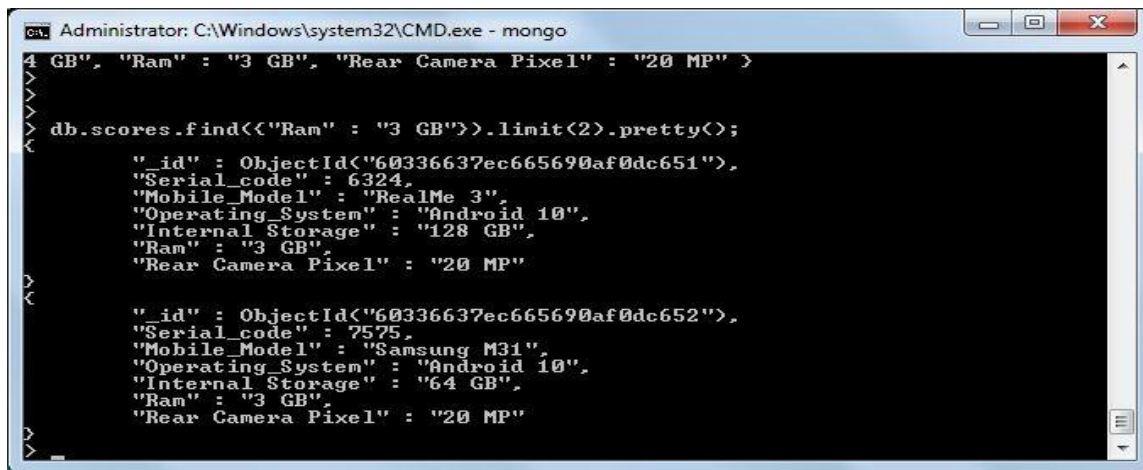


```

{
  "Mobile_Model" : "Asus ch"
}
> db.scores.remove< {"Serial_code" : 8430 } >;
WriteResult<< "nRemoved" : 2 >>
>

```

## 8. Limit command :

Syntax : `db.collection_name.find().limit(number)`

```
Administrator: C:\Windows\system32\CMD.exe - mongo
4 GB", "Ram" : "3 GB", "Rear Camera Pixel" : "20 MP" >
>
> db.scores.find(<<"Ram" : "3 GB">>).limit(2).pretty();
<
  "_id" : ObjectId("60336637ec665690af0dc651"),
  "Serial_code" : 6324,
  "Mobile_Model" : "RealMe 3",
  "Operating_System" : "Android 10",
  "Internal_Storage" : "128 GB",
  "Ram" : "3 GB",
  "Rear_Camera_Pixel" : "20 MP"
>
  "_id" : ObjectId("60336637ec665690af0dc652"),
  "Serial_code" : 7575,
  "Mobile_Model" : "Samsung M31",
  "Operating_System" : "Android 10",
  "Internal_Storage" : "64 GB",
  "Ram" : "3 GB",
  "Rear_Camera_Pixel" : "20 MP"
>
```

## 9. Skip command :

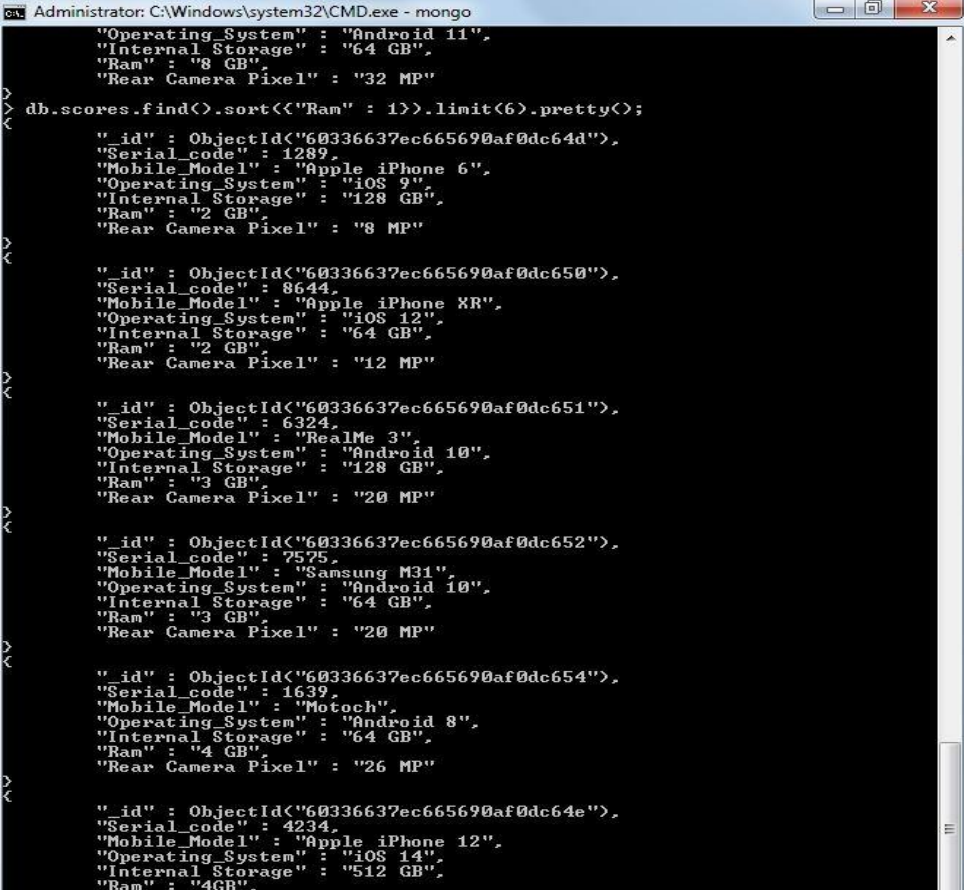
Syntax - `db.collection_name.find().skip(number)`

```
> db.scores.find(<<"Ram" : "3 GB">>).limit(2).skip(1).pretty();
<
  "_id" : ObjectId("60336637ec665690af0dc652"),
  "Serial_code" : 7575,
  "Mobile_Model" : "Samsung M31",
  "Operating_System" : "Android 10",
  "Internal_Storage" : "64 GB",
  "Ram" : "3 GB",
  "Rear_Camera_Pixel" : "20 MP"
>
```

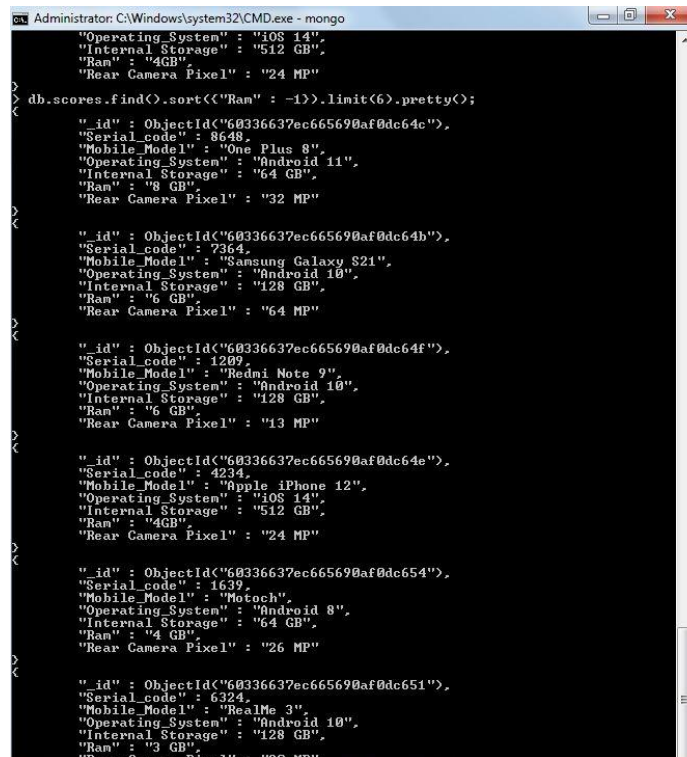
## 10. Sorting commands :

Syntax : **db.collection\_name.find().sort({KEY : 1})**

(Ascending -&gt; 1 , Decending -&gt; -1)

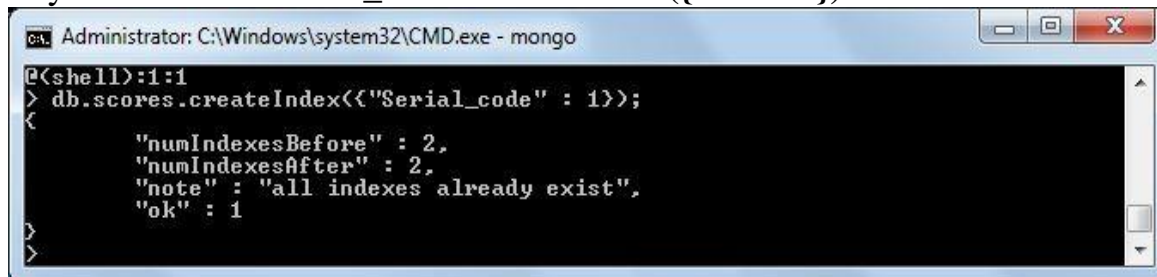


```
Administrator: C:\Windows\system32\CMD.exe - mongo
"Operating_System" : "Android 11",
"Internal Storage" : "64 GB",
"Ram" : "8 GB",
"Rear Camera Pixel" : "32 MP"
> db.scores.find().sort(<"Ram" : 1>>).limit(6).pretty();
<
  "_id" : ObjectId<"60336637ec665690af0dc64d">,
  "Serial_code" : 1289,
  "Mobile_Model" : "Apple iPhone 6",
  "Operating_System" : "iOS 9",
  "Internal Storage" : "128 GB",
  "Ram" : "2 GB",
  "Rear Camera Pixel" : "8 MP"
<
  "_id" : ObjectId<"60336637ec665690af0dc650">,
  "Serial_code" : 8644,
  "Mobile_Model" : "Apple iPhone XR",
  "Operating_System" : "iOS 12",
  "Internal Storage" : "64 GB",
  "Ram" : "2 GB",
  "Rear Camera Pixel" : "12 MP"
<
  "_id" : ObjectId<"60336637ec665690af0dc651">,
  "Serial_code" : 6324,
  "Mobile_Model" : "RealMe 3",
  "Operating_System" : "Android 10",
  "Internal Storage" : "128 GB",
  "Ram" : "3 GB",
  "Rear Camera Pixel" : "20 MP"
<
  "_id" : ObjectId<"60336637ec665690af0dc652">,
  "Serial_code" : 7575,
  "Mobile_Model" : "Samsung M31",
  "Operating_System" : "Android 10",
  "Internal Storage" : "64 GB",
  "Ram" : "3 GB",
  "Rear Camera Pixel" : "20 MP"
<
  "_id" : ObjectId<"60336637ec665690af0dc654">,
  "Serial_code" : 1639,
  "Mobile_Model" : "Motoch",
  "Operating_System" : "Android 8",
  "Internal Storage" : "64 GB",
  "Ram" : "4 GB",
  "Rear Camera Pixel" : "26 MP"
<
  "_id" : ObjectId<"60336637ec665690af0dc64e">,
  "Serial_code" : 4234,
  "Mobile_Model" : "Apple iPhone 12",
  "Operating_System" : "iOS 14",
  "Internal Storage" : "512 GB",
  "Ram" : "4GB",
```



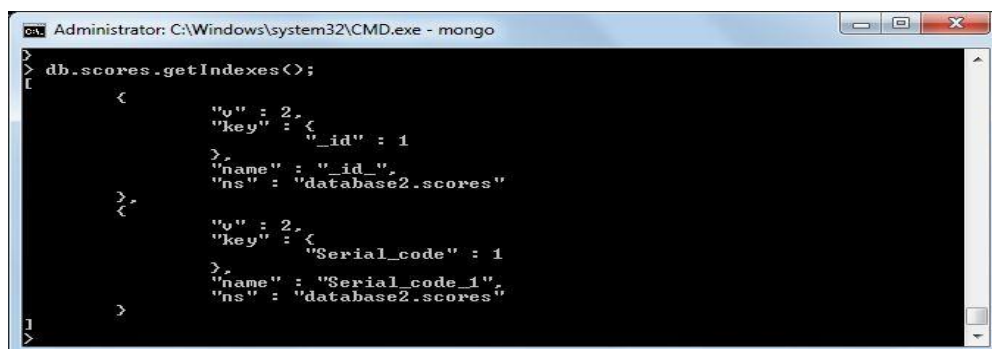
```
Administrator: C:\Windows\system32\CMD.exe - mongo
{"Operating_System": "iOS 14",
 "Internal Storage": "512 GB",
 "Ram": "4GB",
 "Rear Camera Pixel": "24 MP"}
> db.scores.find().sort({"Ram": -1}).limit(6).pretty();
{
  "_id" : ObjectId("60336637ec665690af0dc64c"),
  "Serial_code" : 8648,
  "Mobile_Model" : "One Plus 8",
  "Operating_System": "Android 11",
  "Internal Storage": "64 GB",
  "Ram": "8 GB",
  "Rear Camera Pixel": "32 MP"
}
{
  "_id" : ObjectId("60336637ec665690af0dc64b"),
  "Serial_code" : 7364,
  "Mobile_Model" : "Samsung Galaxy S21",
  "Operating_System": "Android 10",
  "Internal Storage": "128 GB",
  "Ram": "6 GB",
  "Rear Camera Pixel": "64 MP"
}
{
  "_id" : ObjectId("60336637ec665690af0dc64f"),
  "Serial_code" : 1289,
  "Mobile_Model" : "Redmi Note 9",
  "Operating_System": "Android 10",
  "Internal Storage": "128 GB",
  "Ram": "6 GB",
  "Rear Camera Pixel": "13 MP"
}
{
  "_id" : ObjectId("60336637ec665690af0dc64e"),
  "Serial_code" : 4234,
  "Mobile_Model" : "Apple iPhone 12",
  "Operating_System": "iOS 14",
  "Internal Storage": "512 GB",
  "Ram": "4GB",
  "Rear Camera Pixel": "24 MP"
}
{
  "_id" : ObjectId("60336637ec665690af0dc654"),
  "Serial_code" : 1639,
  "Mobile_Model" : "Notoch",
  "Operating_System": "Android 8",
  "Internal Storage": "64 GB",
  "Ram": "4 GB",
  "Rear Camera Pixel": "26 MP"
}
{
  "_id" : ObjectId("60336637ec665690af0dc651"),
  "Serial_code" : 6324,
  "Mobile_Model" : "RealMe 3",
  "Operating_System": "Android 10",
  "Internal Storage": "128 GB",
  "Ram": "3 GB",
  "Rear Camera Pixel": "20 MP"
}
```

## 11. Create Index command

Syntax : **db.collection\_name.createIndex({KEY :1})**


```
Administrator: C:\Windows\system32\CMD.exe - mongo
@(<shell>):1:1
> db.scores.createIndex({"Serial_code" : 1});
{
  "numIndexesBefore" : 2,
  "numIndexesAfter" : 2,
  "note" : "all indexes already exist",
  "ok" : 1
}
```

## 12. Get Index command :

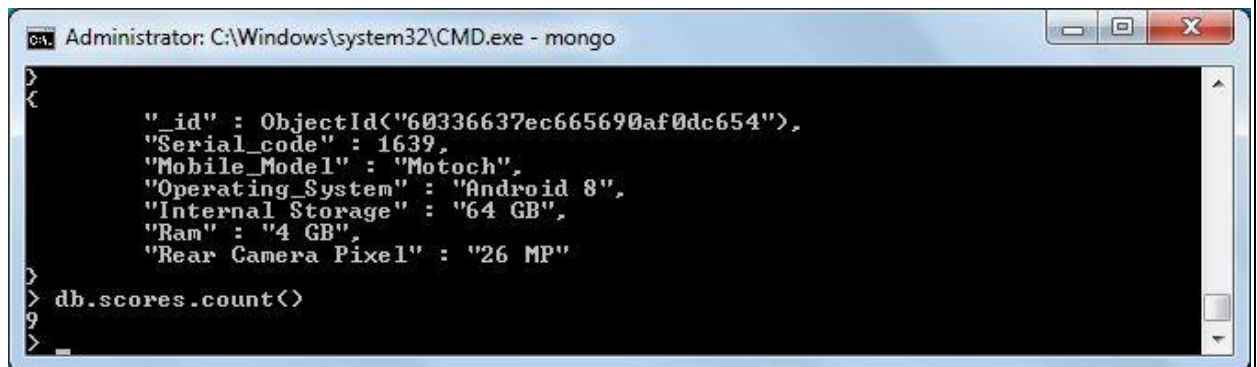
Syntax : **db.collection\_name.getIndexes()**


```
Administrator: C:\Windows\system32\CMD.exe - mongo
> db.scores.getIndexes();
[
  {
    "u" : 2,
    "key" : {
      "_id" : 1
    },
    "name" : "_id_",
    "ns" : "database2.scores"
  },
  {
    "u" : 2,
    "key" : {
      "Serial_code" : 1
    },
    "name" : "Serial_code_1",
    "ns" : "database2.scores"
  }
]
```



13.Count command :

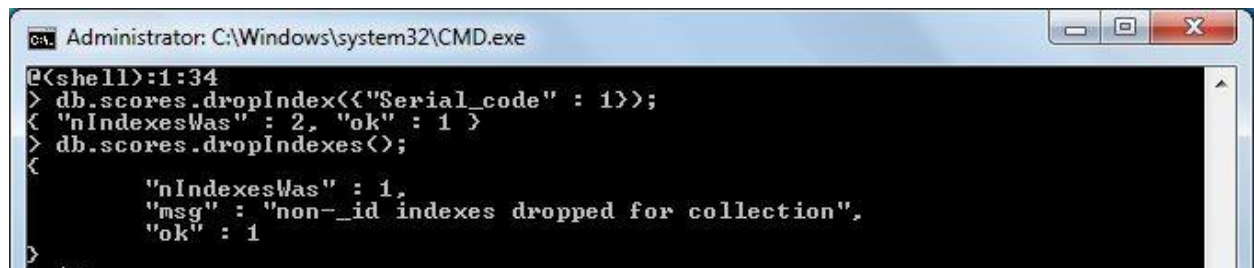
Syntax : **db.collection\_name.count()**



```
Administrator: C:\Windows\system32\CMD.exe - mongo
>
<
  "_id" : ObjectId<"60336637ec665690af0dc654">,
  "Serial_code" : 1639,
  "Mobile_Model" : "Motoch",
  "Operating_System" : "Android 8",
  "Internal Storage" : "64 GB",
  "Ram" : "4 GB",
  "Rear Camera Pixel" : "26 MP"
>
> db.scores.count()
1
>
```

14.Drop indexes command :

Syntax : **db.collection\_name.dropIndexes()**



```
Administrator: C:\Windows\system32\CMD.exe
@<shell>:1:34
> db.scores.dropIndex(<<"Serial_code" : 1>>);
< "nIndexesWas" : 2, "ok" : 1 }
> db.scores.dropIndexes();
<
  "nIndexesWas" : 1,
  "msg" : "non-_id indexes dropped for collection",
  "ok" : 1
>
```