Roll No : 757 **Page No : ___**

Date : 22/ 02 /2021

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.

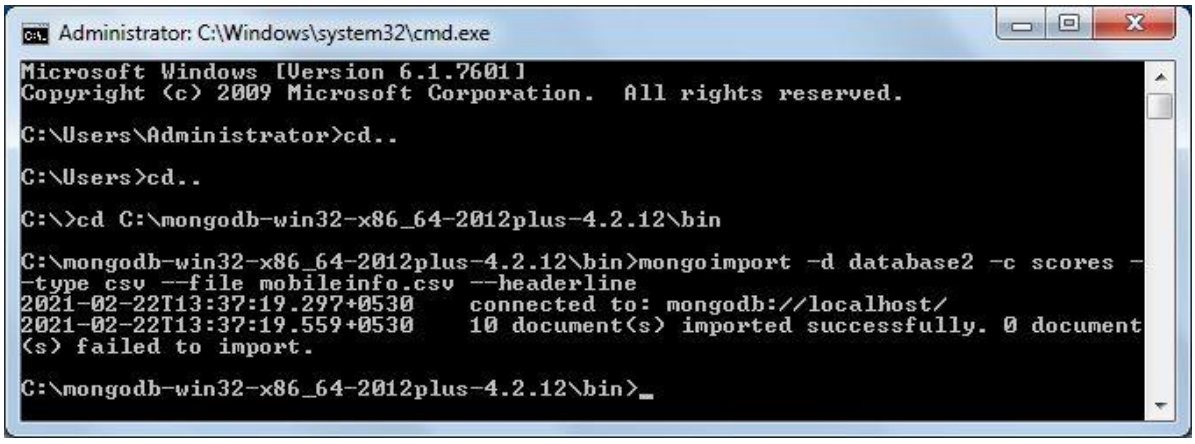
Data (csv format) :

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

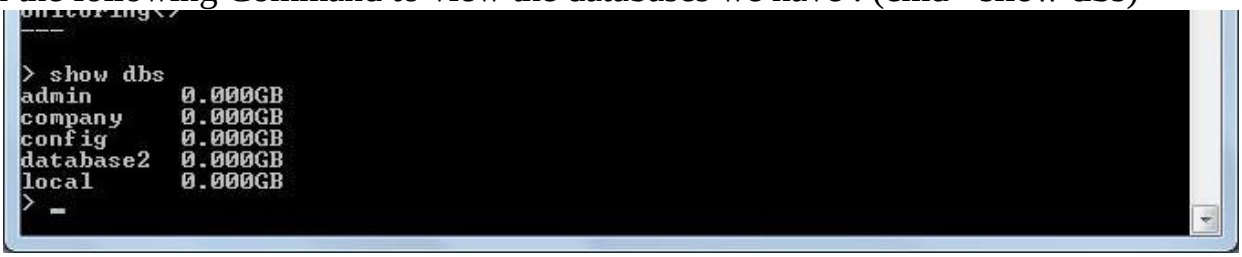
Steps :

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**

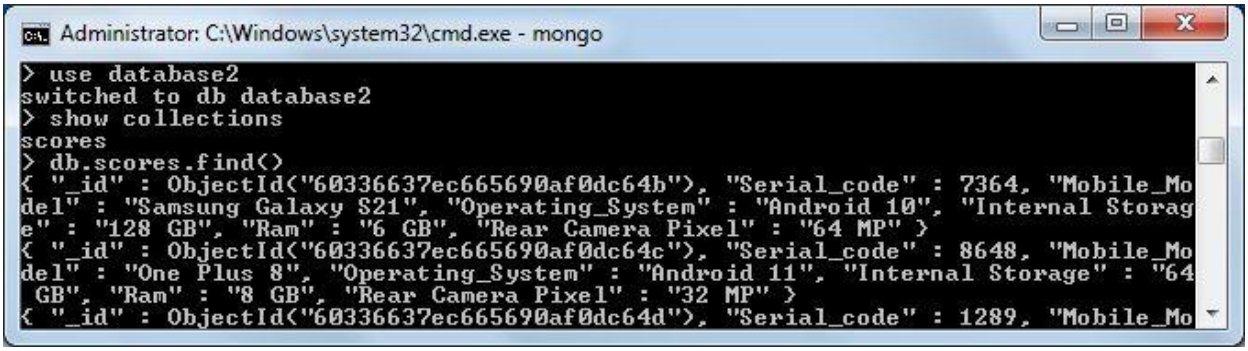


2. Run the following Command to view the databases we have : (cmd - **show dbs**)



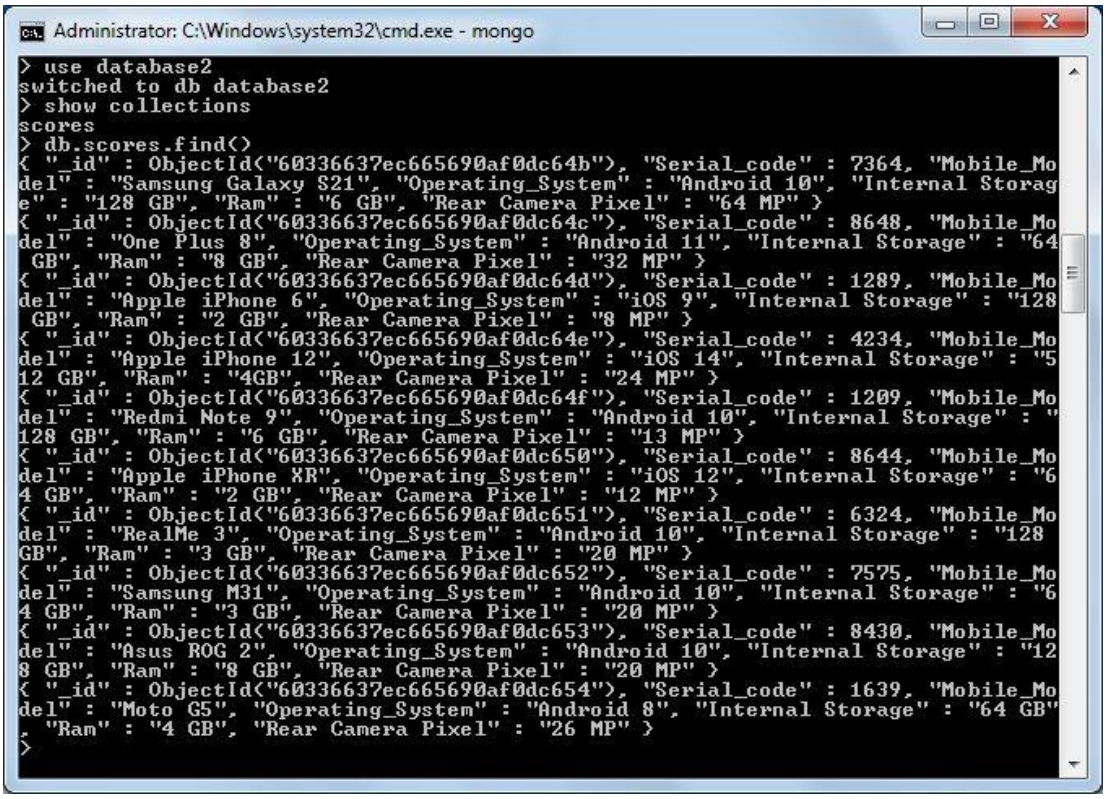
3.Now we can use our database :

Syntax : **use database_name**



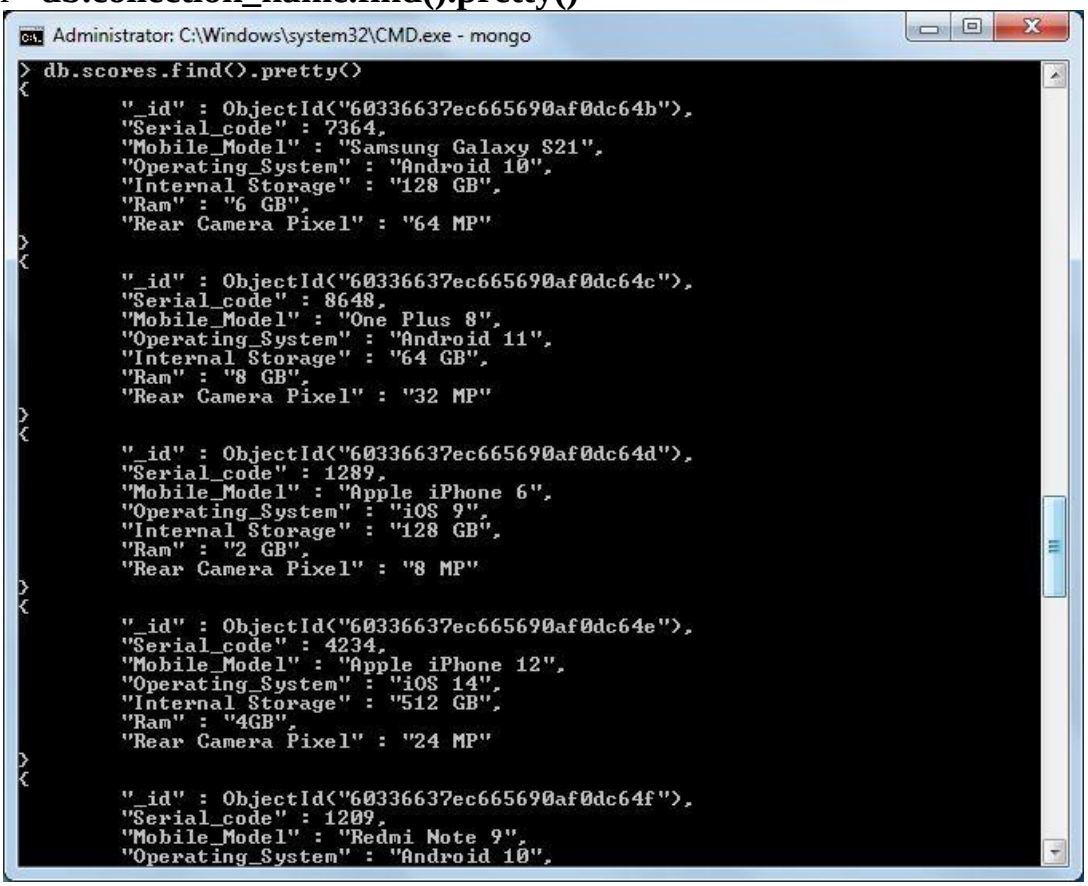
4. Now we have to get the collections we have in our database for that we have to run the following command :

Syntax : **show collections**



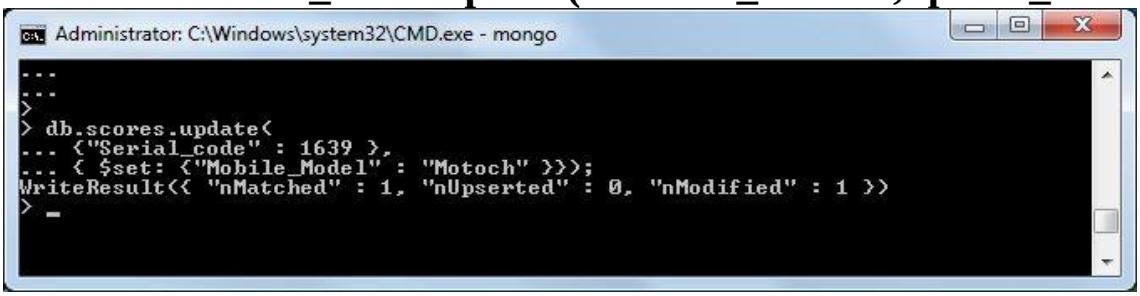
5. For extracting the documents from collections we have the following command :

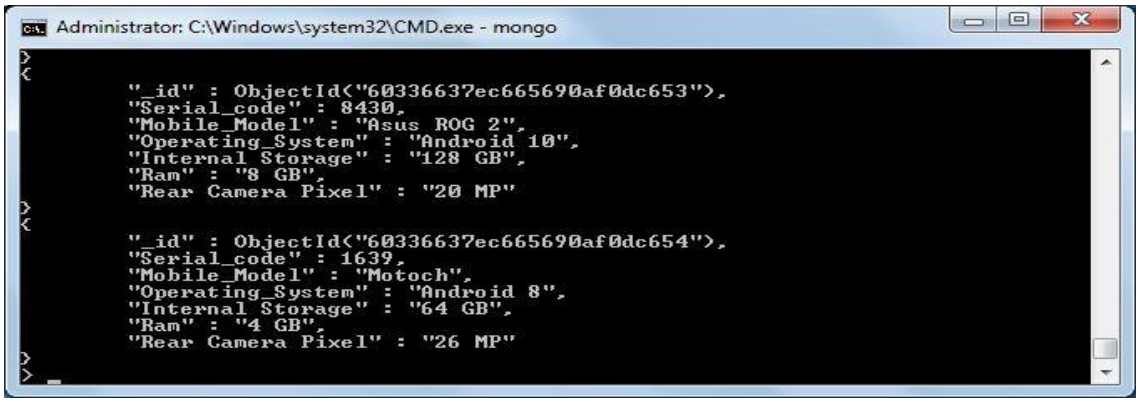
Syntax - **db.collection_name.find().pretty()**

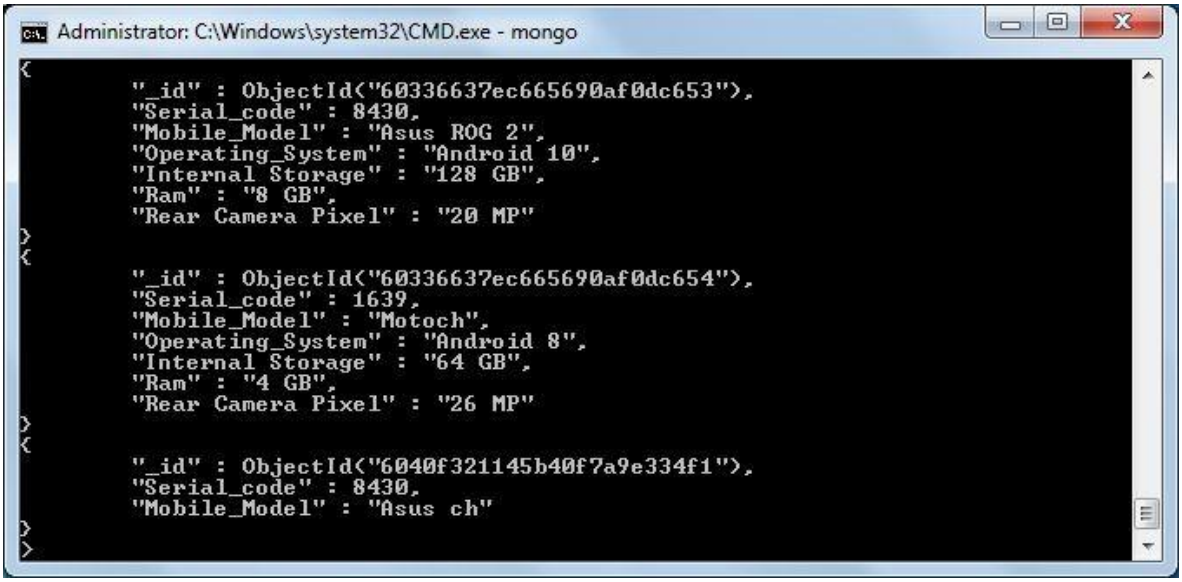


6. Now we can perform basic mongodb commands to manipulate the data in our database. The command we would be using first is Update command .

Syntax – **db.collection_name.update(selection_criteria,update_criteria)**





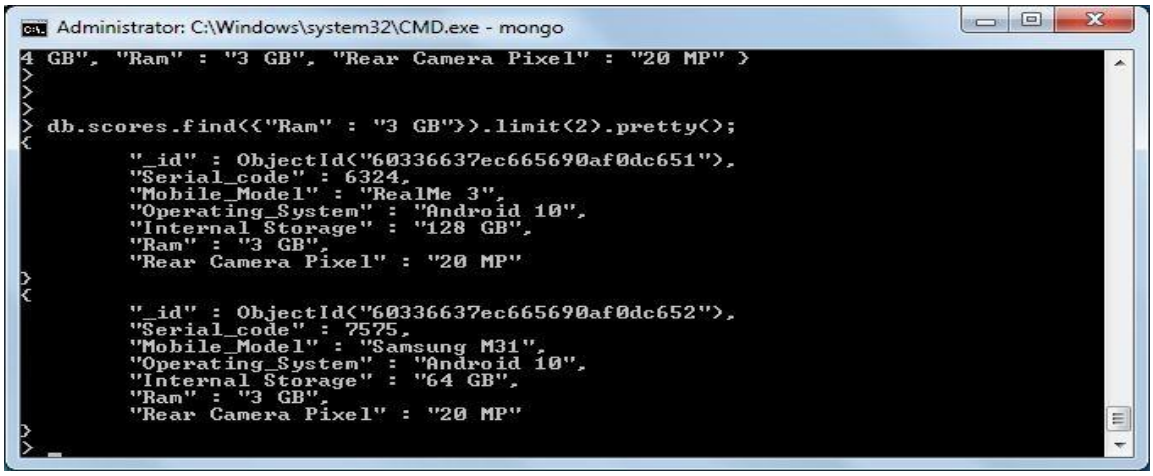


7.Next command we will be using is remove command :

Syntax : **db.collection_name.remove(deletion_criteria**)

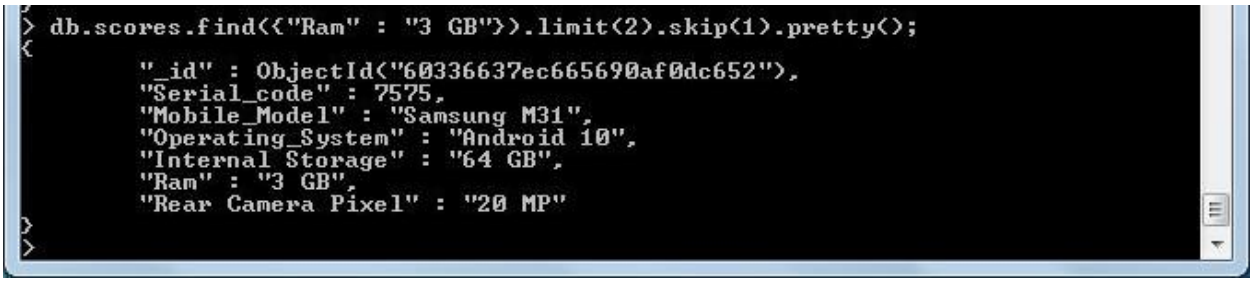
8.Limit command :

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



9. Skip command :

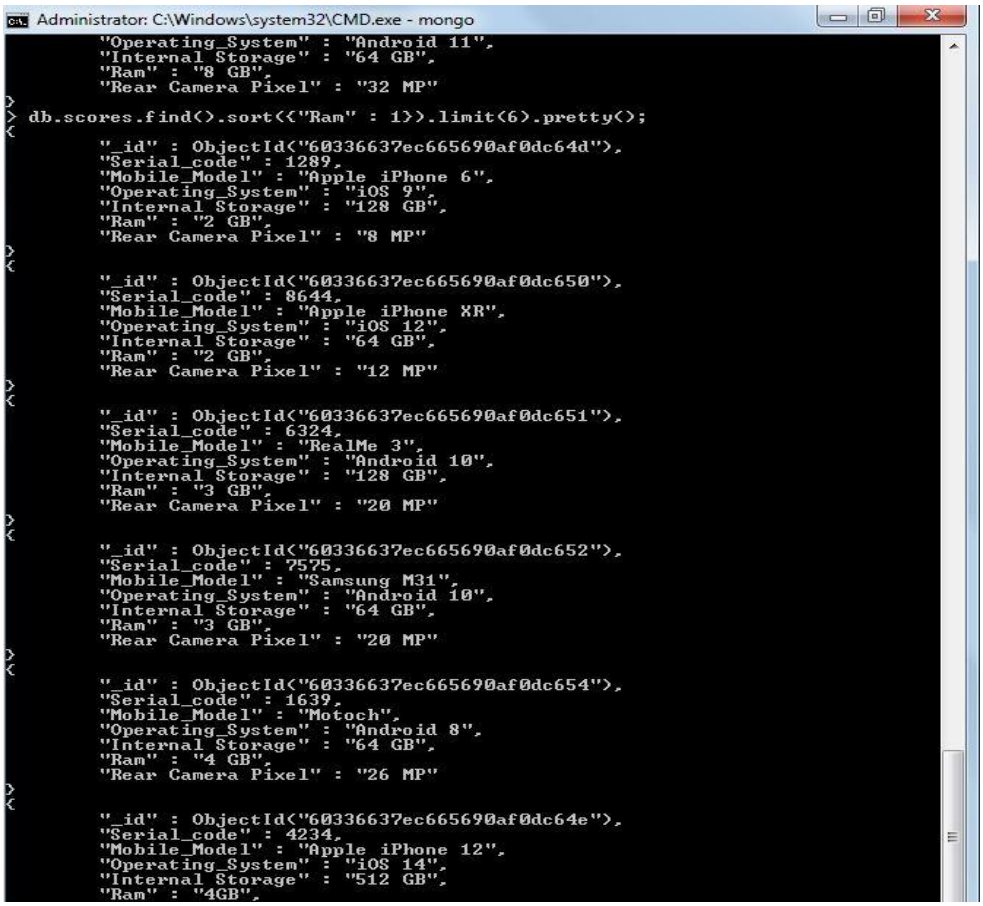
Syntax - **db.collection_name.find().skip(number)**

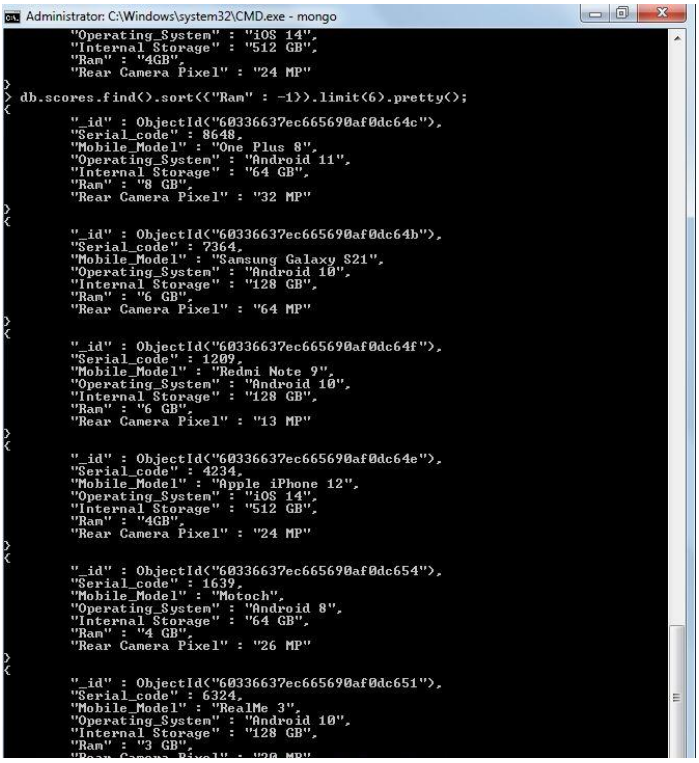


10. Sorting commands :

Syntax : **db.collection_name.find().sort({KEY : 1})**

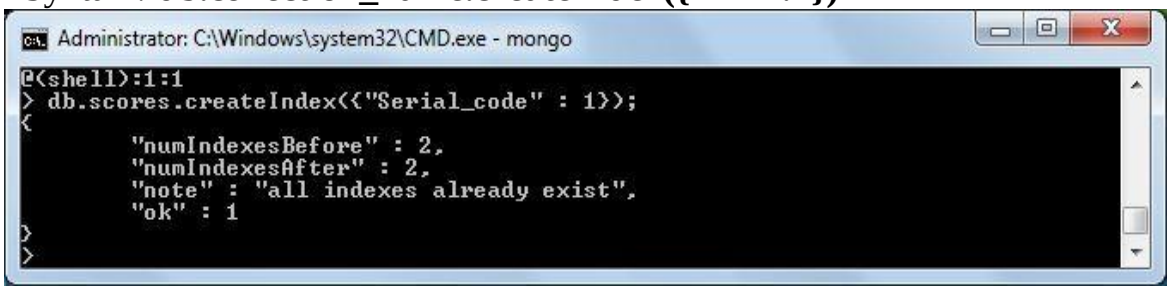
(Ascending -> 1 , Decending -> -1)





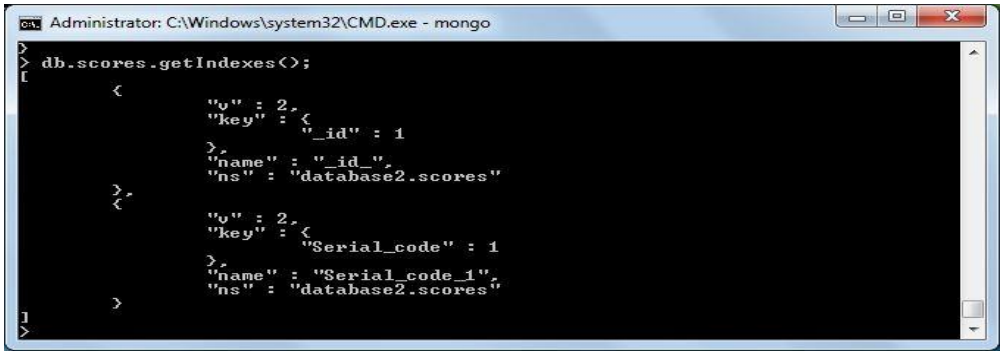
11. Create Index command

Syntax : **db.collection_name.createIndex({KEY :1})**



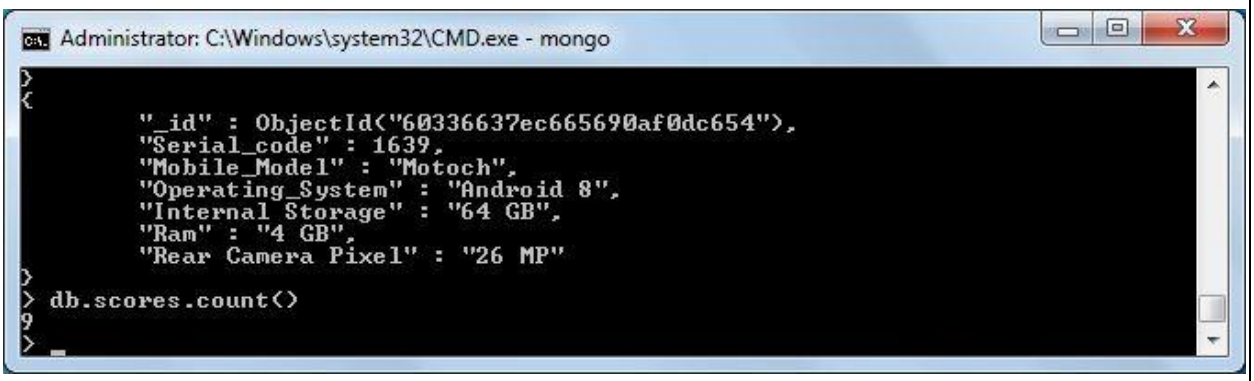
12. Get Index command :

Syntax : **db.collection_name.getIndexes()**



13.Count command :

Syntax : **db.collection_name.count()**



14.Drop indexes command :

Syntax : **db.collection_name.dropIndexes()**