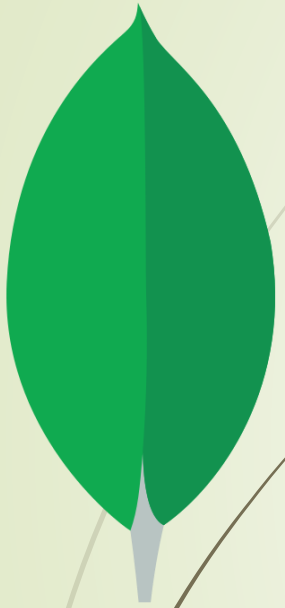


mongoDB®

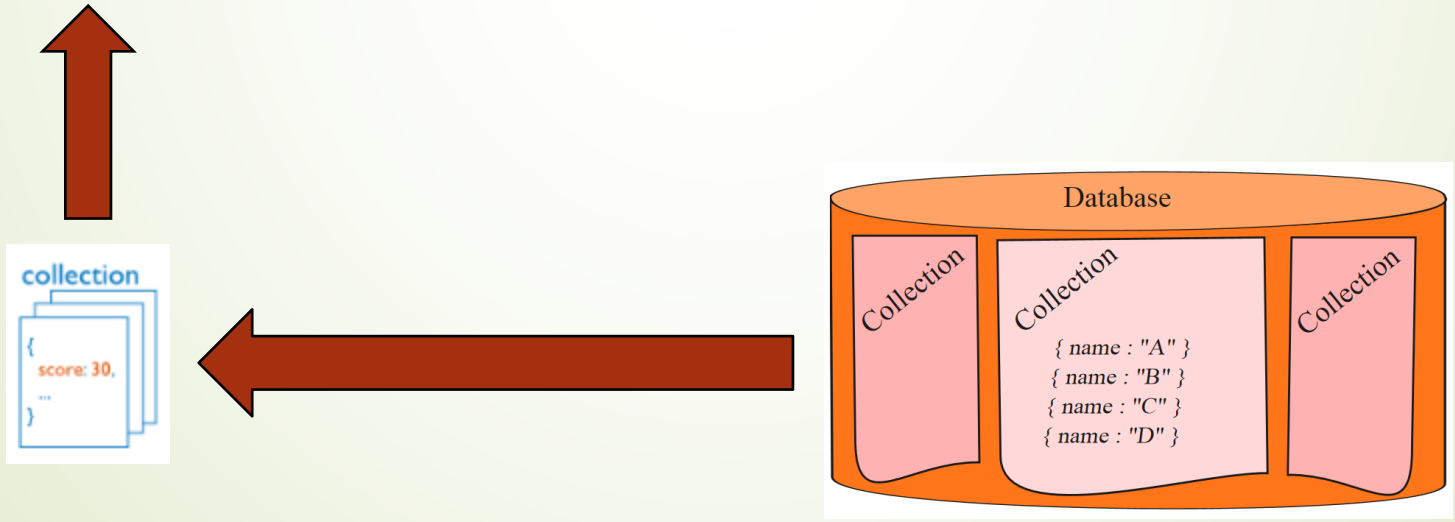
Bigdata Lab 7

MongoDB

Before Index

- What does database normally do when we query?
  - MongoDB must scan every document.
  - Inefficient because process large volume of data

db.users.find({ score: { “$lt” : 30} })



Definition of Index

- **Definition**
  - **Indexes are special data structures that store a small portion of the collection's data set in an easy to traverse form.**

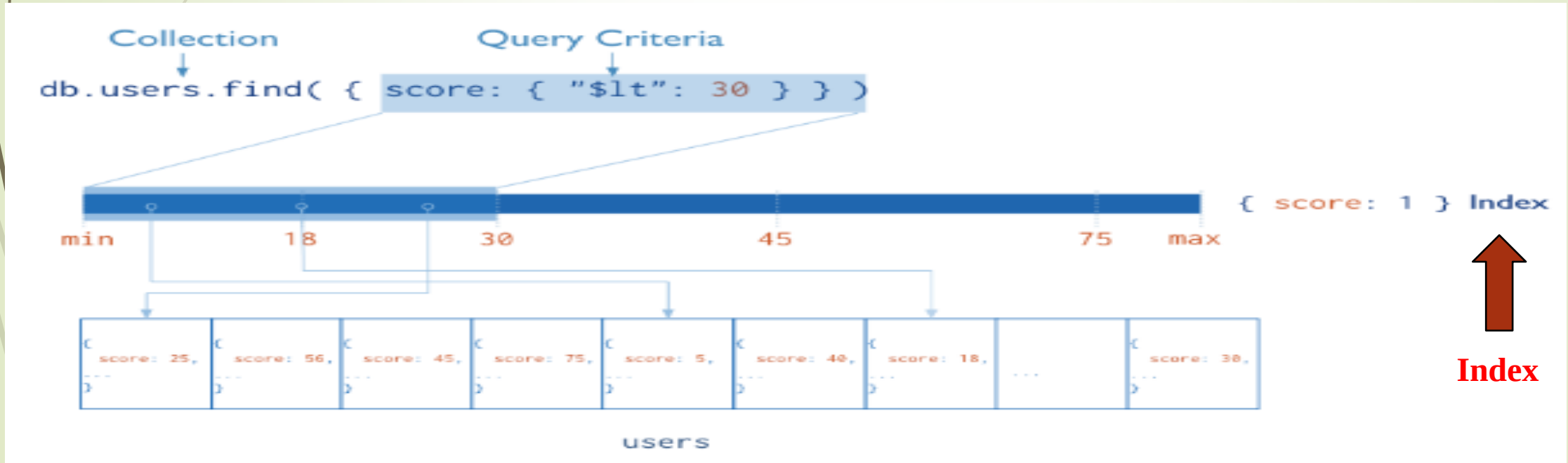


Diagram of a query that uses an index to select

Index in MongoDB

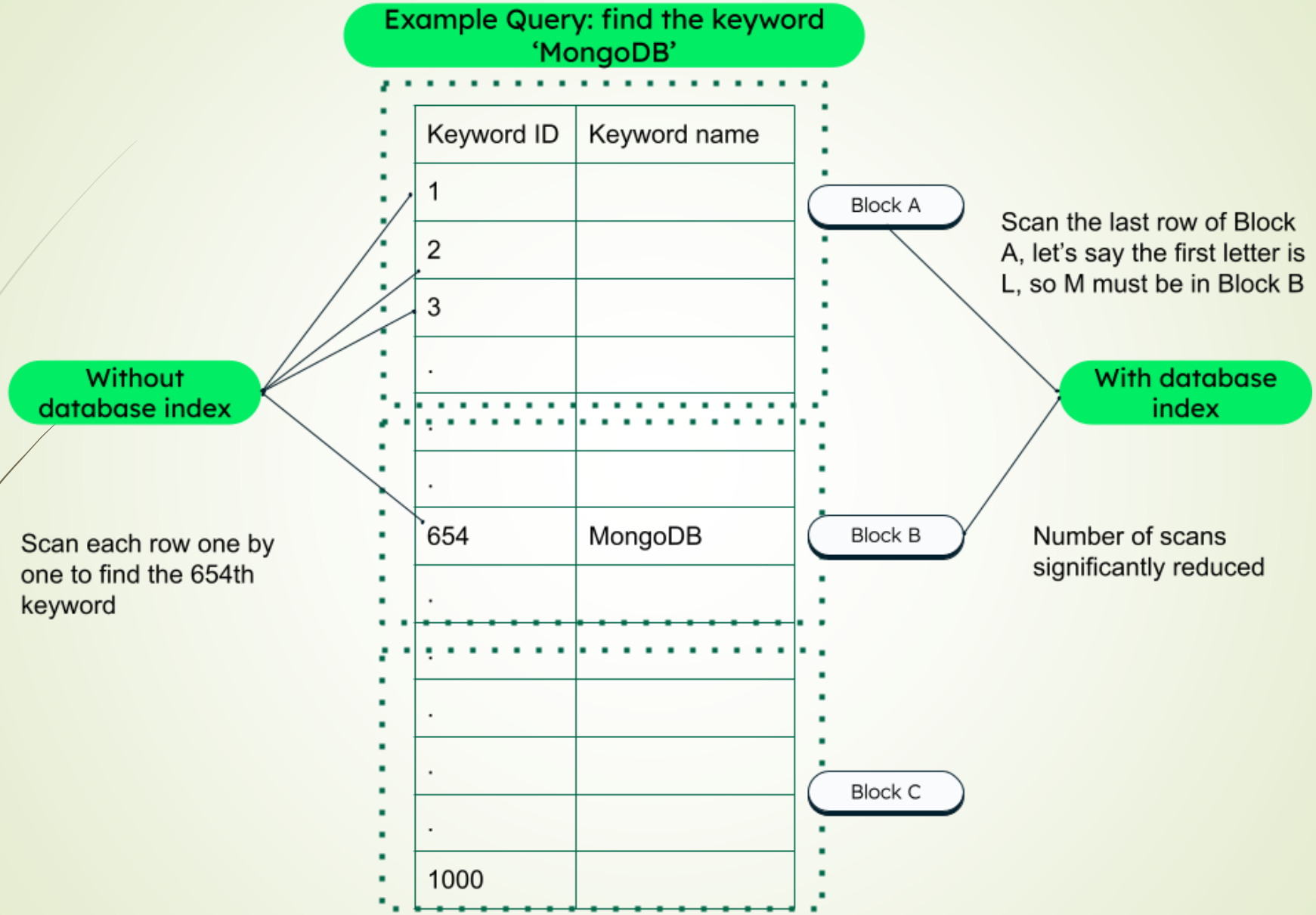
Operations

- **Creation index**
  - db.users.ensureIndex({ score: 1 })

- By default each created collections mongo create an index base on **_id** and sorted ascending**.**
- **Show existing indexes**
  - db.users.getIndexes()
- **Drop index**
  - db.users.dropIndex({score: 1})

Index in MongoDB

- **Explain—Explain**
  - db.users.find().explain()
  - Returns a document that **describes the process and indexes and ensure that the find() function used the index in the query.**
- **Hint**
  - db.users.find().hint({score: 1})
  - Overide MongoDB’s **default index selection**


Example Query: find the keyword 'MongoDB'
Keyword ID
Keyword name
1
2
3
654
MongoDB
1000
Block A
Block B
Block C
Without database index
With database index
Scan each row one by one to find the 654th keyword
Scan the last row of Block A, let's say the first letter is L, so M must be in Block B
Number of scans significantly reduced

Indexing Properties in mongoDB

- **background: Boolean**
  - Builds the index in the background.(default: false)
- **unique: Boolean**
  - Reject duplicate values in the index field.(Default: false)
- **name: String**
  - Change the index name.(Default : attributeName:1 Or -1)
- **sparse: Boolean**
  - Skip any document that miss the index key
- **expireAfterSeconds**:

Index in MongoDB

- Types of Indices:
  - Single Field Indexes
  - Compound Field Indexes
  - Multikey Indexes

Single Field Index

- **db.users.ensureIndex({ score: 1 })**

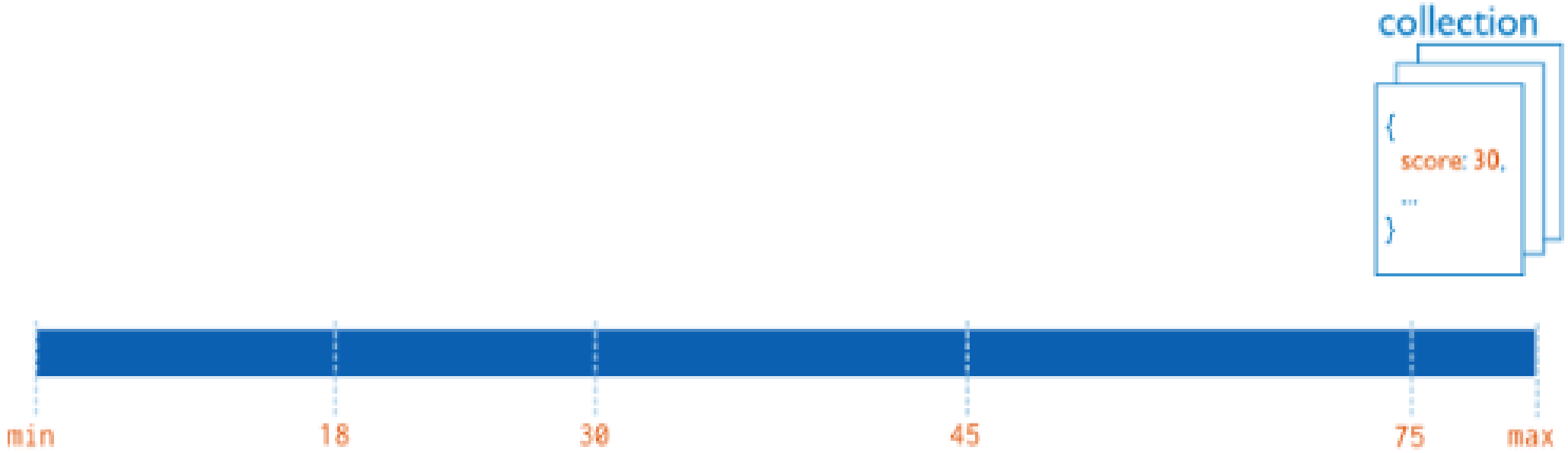


Diagram of an index on the score field (ascending).

Compound Field index

- **db.users.ensureIndex({ userid:1, score: -1 })**

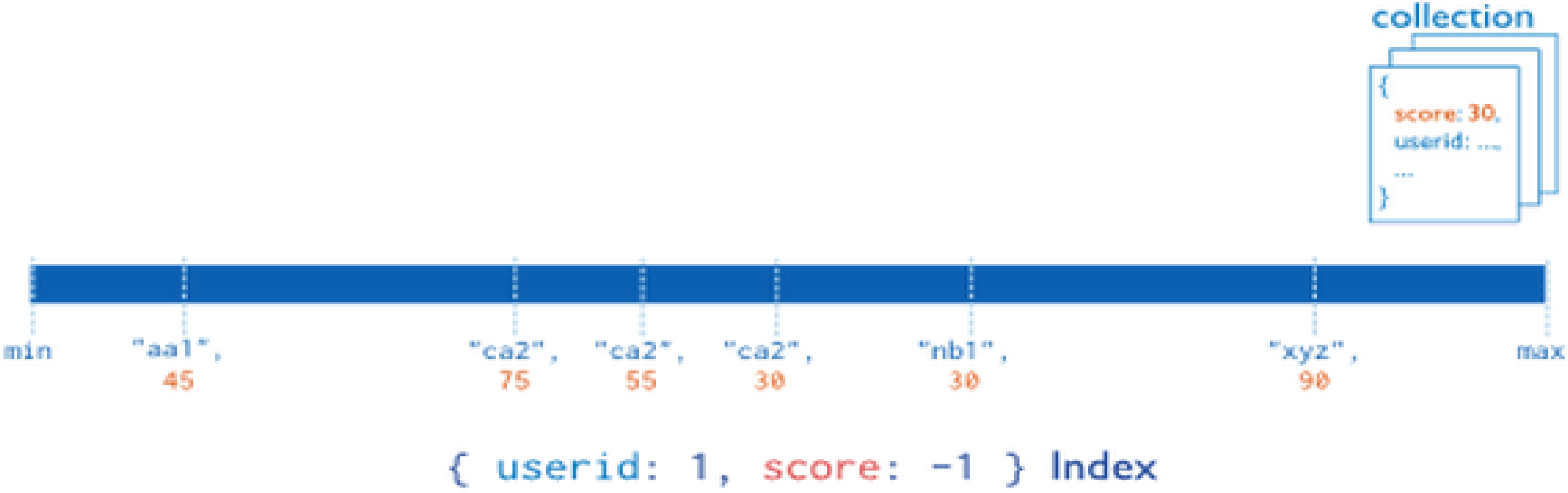


Diagram of a compound index on the `userid` field (ascending) and the `score` field (descending). The index sorts first by the `userid` field and then by the `score` field.

Multi-Key index

- **db.users.ensureIndex({ addr.zip:1})**

Diagram of a multikey index on the addr.zip field. The addr field contains an array of address documents. The address documents contain the zip field.

Let's try

Go to commands file

Thanks

Reference:
A. Im, G. Cai, H. Tunc, J. Stevens, Y. Barve, S. Hei
Vanderbilt University