Firdaus Mulla | L007

Mithibai College Department of Computer Science

Msc(Data Science and AI)

Practical 4: INDEXING USING Mongodb

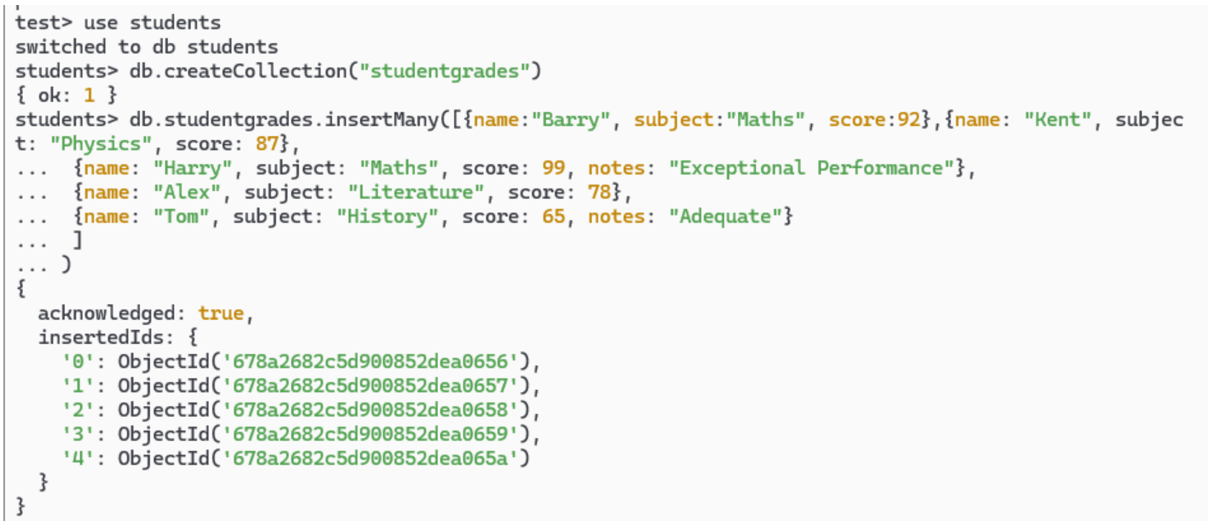
Date: 17/01/2024 **Submission Date: 24/01/2024**

Write-up: -

- Indexing in mongodb
- Types of indexing
- Document document-oriented NoSQL

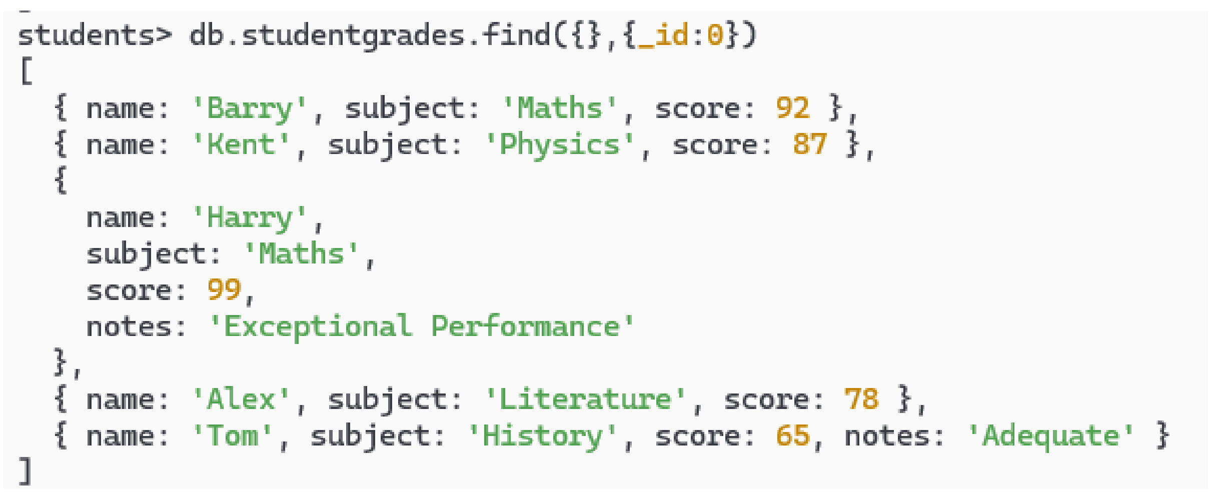
1. Mongo DB indexing
a. Create index in Mongo DB
b. Finding the indexes in a collection
c. Drop indexes in a collection
d. Drop all the indexes

```
use students
db.createCollection("studentgrades")
db.studentgrades.insertMany(
 [
 {name: "Barry", subject: "Maths", score: 92},
 {name: "Kent", subject: "Physics", score: 87},
 {name: "Harry", subject: "Maths", score: 99, notes: "Exceptional Performance"},
 {name: "Alex", subject: "Literature", score: 78},
 {name: "Tom", subject: "History", score: 65, notes: "Adequate"}
 ]
)
```



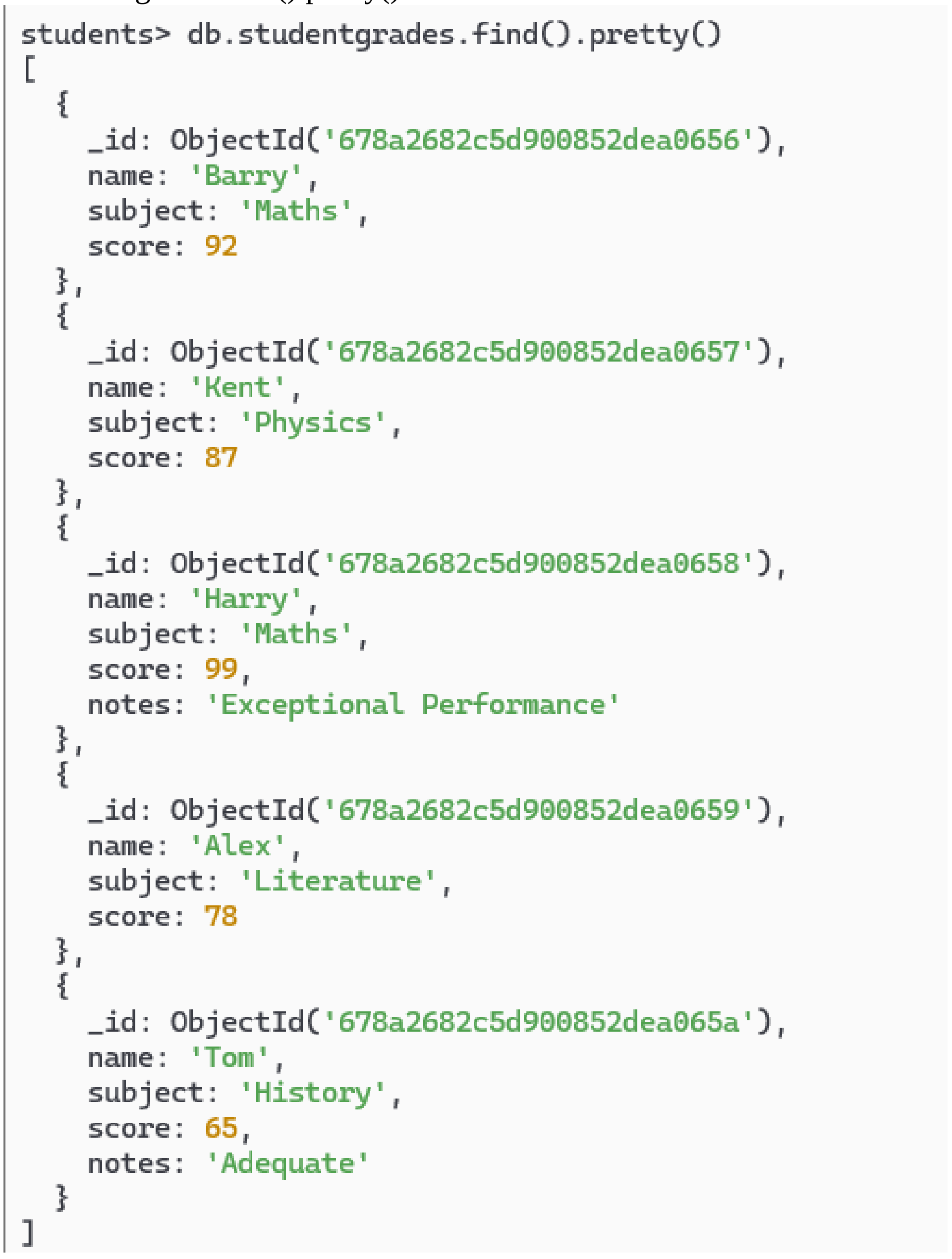
```
test> use students
switched to db students
students> db.createCollection("studentgrades")
{ ok: 1 }
students> db.studentgrades.insertMany([{name:"Barry", subject:"Maths", score:92},{name: "Kent", subjec
t: "Physics", score: 87},
...   {name: "Harry", subject: "Maths", score: 99, notes: "Exceptional Performance"},
...   {name: "Alex", subject: "Literature", score: 78},
...   {name: "Tom", subject: "History", score: 65, notes: "Adequate"}
...   ]
... )
{
  acknowledged: true,
  insertedIds: {
    '0': ObjectId('678a2682c5d900852dea0656'),
    '1': ObjectId('678a2682c5d900852dea0657'),
    '2': ObjectId('678a2682c5d900852dea0658'),
    '3': ObjectId('678a2682c5d900852dea0659'),
    '4': ObjectId('678a2682c5d900852dea065a')
  }
}
```

```
db.studentgrades.find({},{_id:0})
```



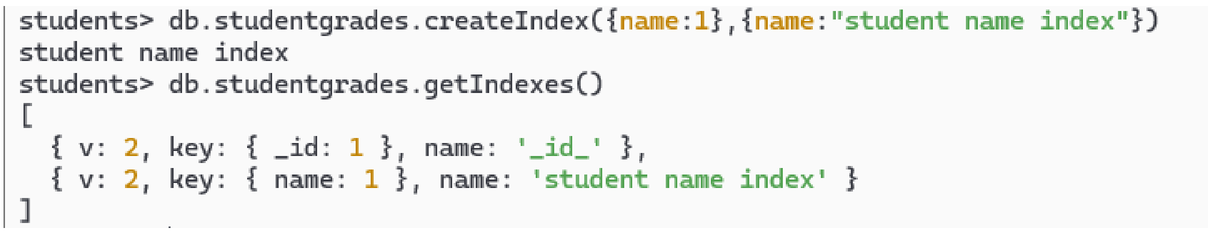
```
students> db.studentgrades.find({},{_id:0})
[
  { name: 'Barry', subject: 'Maths', score: 92 },
  { name: 'Kent', subject: 'Physics', score: 87 },
  {
    name: 'Harry',
    subject: 'Maths',
    score: 99,
    notes: 'Exceptional Performance'
  },
  { name: 'Alex', subject: 'Literature', score: 78 },
  { name: 'Tom', subject: 'History', score: 65, notes: 'Adequate' }
]
```

db.studentgrades.find().pretty()



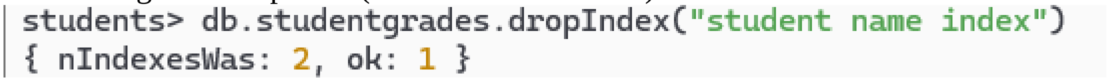
```
students> db.studentgrades.find().pretty()
[
  {
    _id: ObjectId('678a2682c5d900852dea0656'),
    name: 'Barry',
    subject: 'Maths',
    score: 92
  },
  {
    _id: ObjectId('678a2682c5d900852dea0657'),
    name: 'Kent',
    subject: 'Physics',
    score: 87
  },
  {
    _id: ObjectId('678a2682c5d900852dea0658'),
    name: 'Harry',
    subject: 'Maths',
    score: 99,
    notes: 'Exceptional Performance'
  },
  {
    _id: ObjectId('678a2682c5d900852dea0659'),
    name: 'Alex',
    subject: 'Literature',
    score: 78
  },
  {
    _id: ObjectId('678a2682c5d900852dea065a'),
    name: 'Tom',
    subject: 'History',
    score: 65,
    notes: 'Adequate'
  }
]
```

db.studentgrades.createIndex({name: 1}, {name: "student name index"})
Finding indexes You can find all the available indexes in a MongoDB collection by using the getIndexes method. This will return all the indexes in a specific collection. db..getIndexes() Let's view all the indexes in the studentgrades collection using the following command:
db.studentgrades.getIndexes()



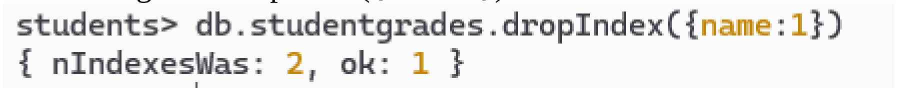
```
students> db.studentgrades.createIndex({name:1},{name:"student name index"})
student name index
students> db.studentgrades.getIndexes()
[
  { v: 2, key: { _id: 1 }, name: '_id_' },
  { v: 2, key: { name: 1 }, name: 'student name index' }
]
```

Dropping indexes To delete an index from a collection, use the dropIndex method while specifying the index name to be dropped. db..dropIndex() Let's remove the user-created index with the index name student name index, as shown below. db.studentgrades.dropIndex("student name index")



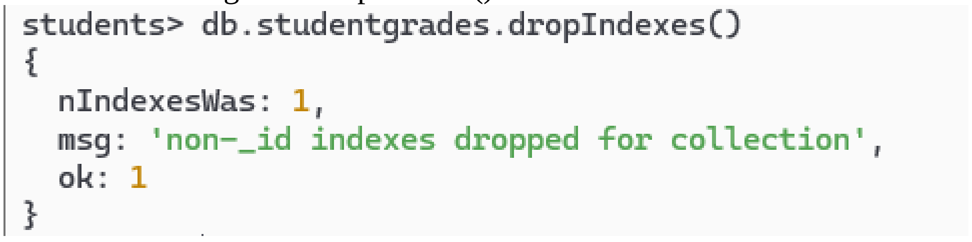
```
students> db.studentgrades.dropIndex("student name index")
{ nIndexesWas: 2, ok: 1 }
```

You can also use the index field value for removing an index without a defined name: db.studentgrades.dropIndex({name:1})



```
students> db.studentgrades.dropIndex({name:1})
{ nIndexesWas: 2, ok: 1 }
```

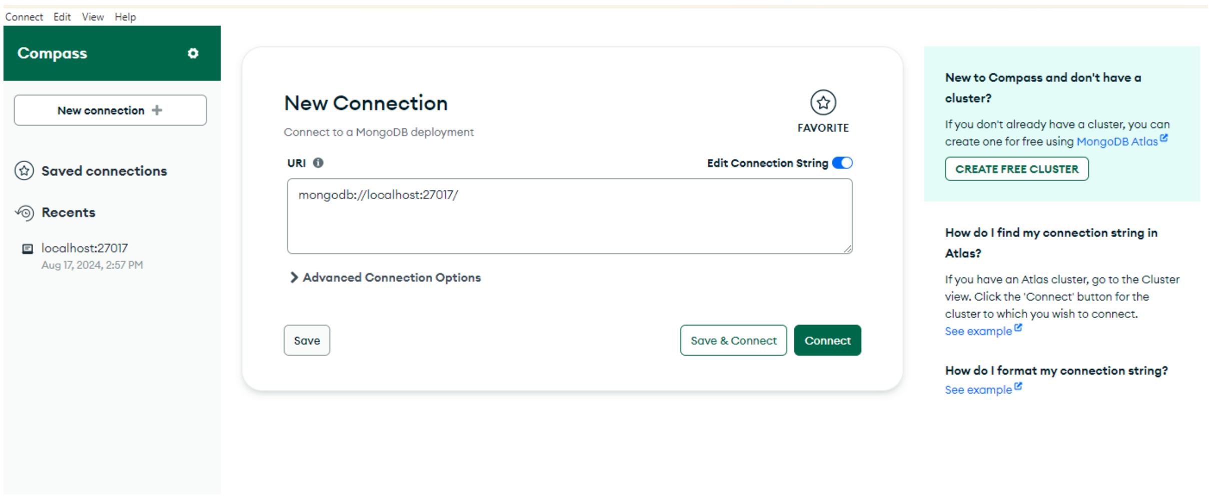
The dropIndexes command can also drop all the indexes excluding the default _id index. db.studentgrades.dropIndexes()



```
students> db.studentgrades.dropIndexes()
{
  nIndexesWas: 1,
  msg: 'non-_id indexes dropped for collection',
  ok: 1
}
```

2. Create all the types of indexes (discussed in class) which will help in finding certain words in a document by using AIRPORT (dataset).
Create connection

Firdaus Mulla | L007


Connect Edit View Help

Compass

New connection +

Saved connections

Recents

localhost:27017
Aug 17, 2024, 2:57 PM

New Connection

Connect to a MongoDB deployment

FAVORITE

URI

Edit Connection String

```
mongodb://localhost:27017/
```

Advanced Connection Options

Save | Save & Connect | Connect

New to Compass and don't have a cluster?

If you don't already have a cluster, you can create one for free using MongoDB Atlas

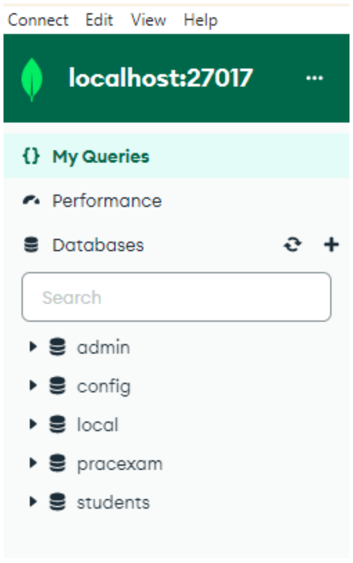
CREATE FREE CLUSTER

How do I find my connection string in Atlas?

If you have an Atlas cluster, go to the Cluster view. Click the 'Connect' button for the cluster to which you wish to connect.
See example

How do I format my connection string?
See example

Create a New Database


Connect Edit View Help

localhost:27017

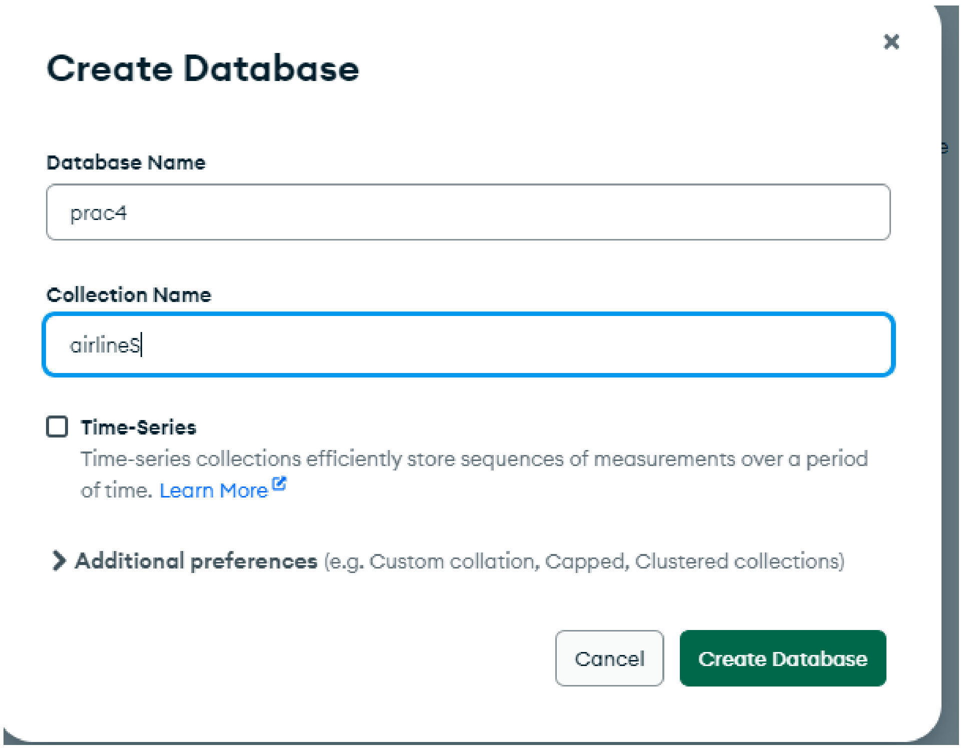
My Queries

Performance

Databases

Search

- admin
- config
- local
- pracexam
- students


Create Database

Database Name: prac4

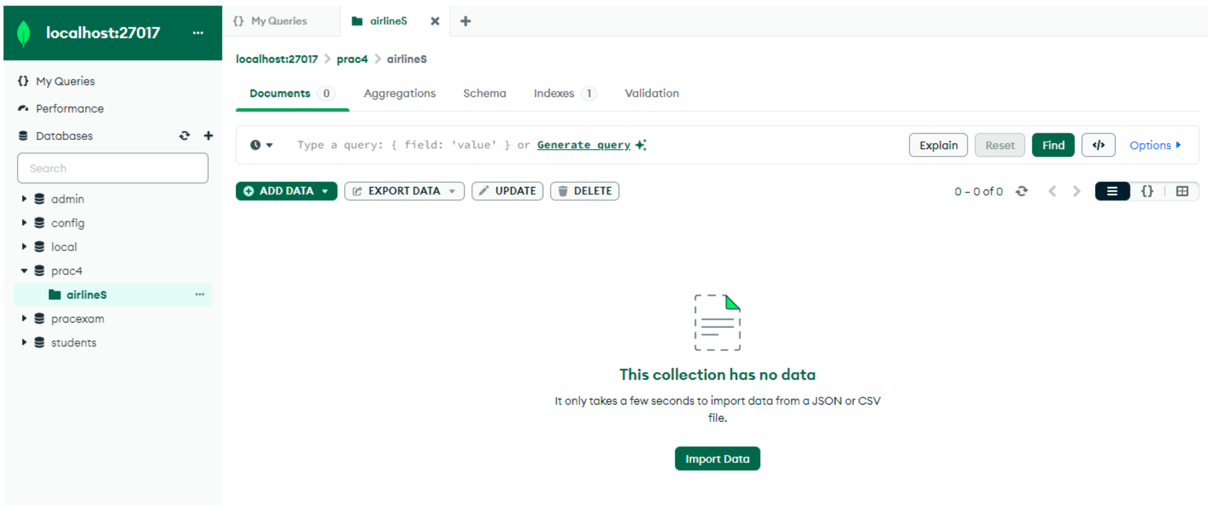
Collection Name: airlineS

Time-Series
Time-series collections efficiently store sequences of measurements over a period of time. Learn More

Additional preferences (e.g. Custom collation, Capped, Clustered collections)

Cancel | Create Database

Firdaus Mulla | L007


localhost:27017 > prac4 > airlineS

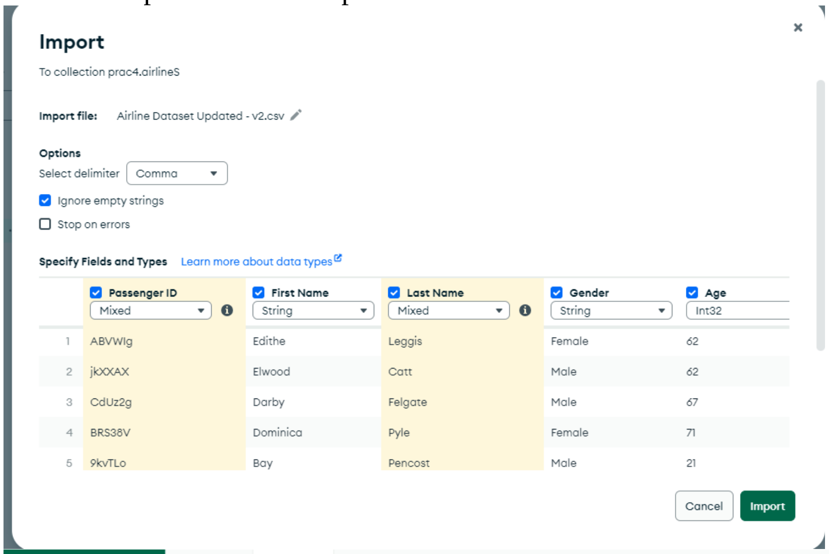
Documents 0 | Aggregations | Schema | Indexes 1 | Validation

Type a query: { field: 'value' } or Generate query

This collection has no data

It only takes a few seconds to import data from a JSON or CSV file.

Click on Import Data and import the csv file


Import

To collection prac4.airlineS

Import file: Airline Dataset Updated - v2.csv

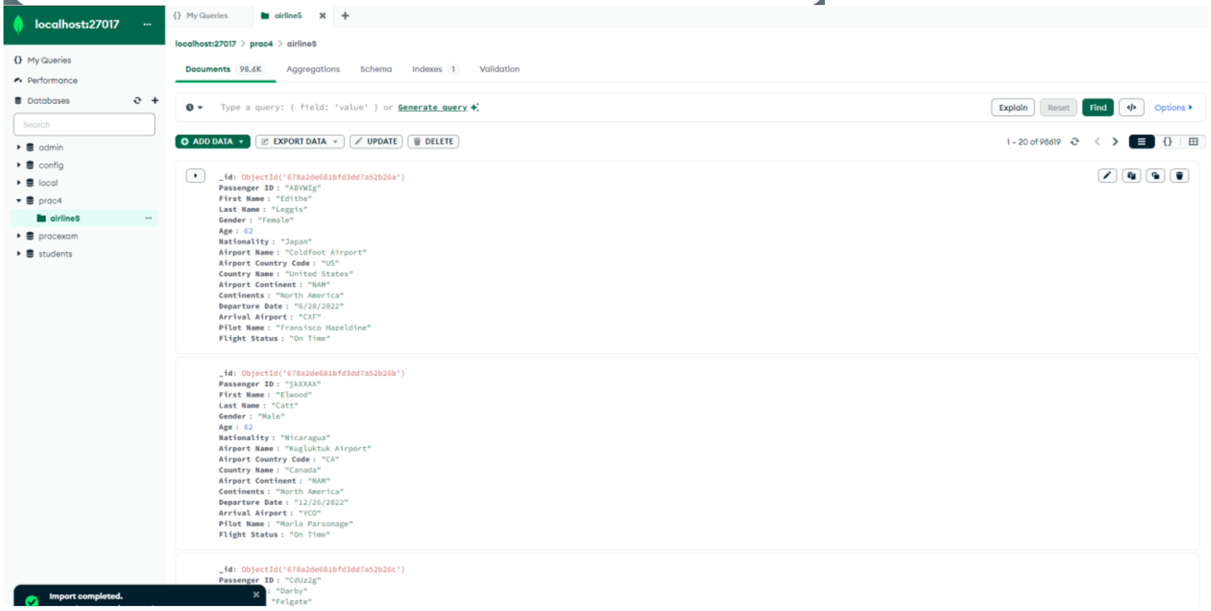
Options

Select delimiter: Comma

- [x] Ignore empty strings
- [ ] Stop on errors

Specify Fields and Types

	☑ Passenger ID (Mixed)	☑ First Name (String)	☑ Last Name (Mixed)	☑ Gender (String)	☑ Age (Int32)
1	ABVWIg	Edithe	Leggis	Female	62
2	jkXXAX	Elwood	Catt	Male	62
3	CdUz2g	Darby	Felgate	Male	67
4	BRS38V	Dominica	Pyle	Female	71
5	9kvTLo	Bay	Pencost	Male	21


localhost:27017 > prac4 > airlineS

Documents 98.6K | Aggregations | Schema | Indexes 1 | Validation

1 - 20 of 98619

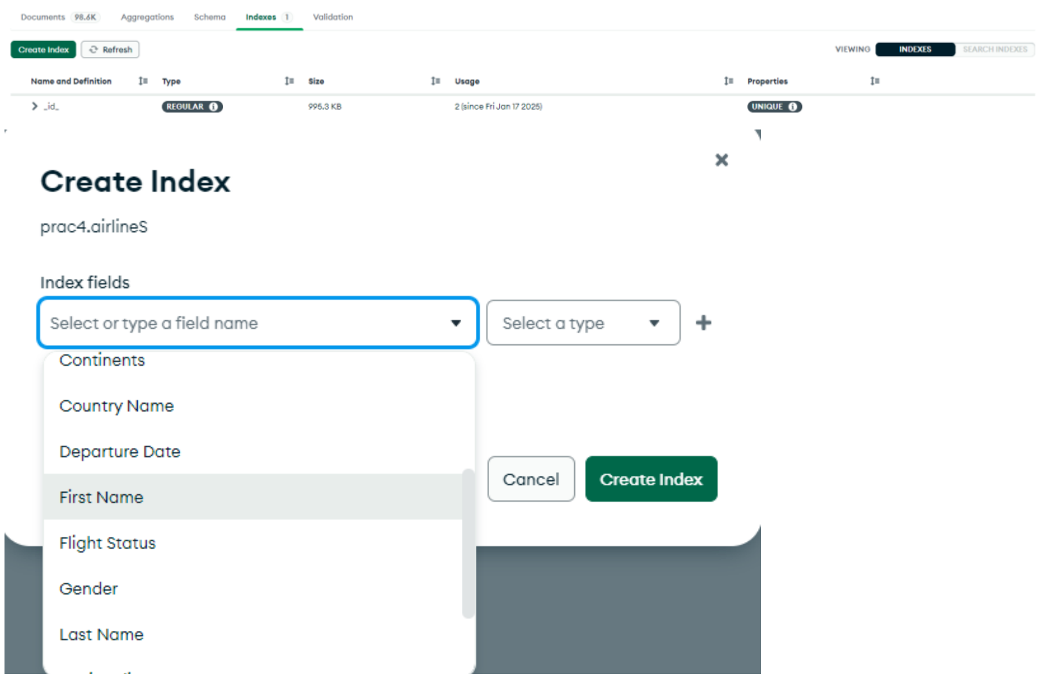
```
_id: ObjectId('678a2de681bfd3dd7a52b26a')
Passenger ID : "ABVWIg"
First Name : "Edithe"
Last Name : "Leggis"
Gender : "Female"
Age : 62
Nationality : "Japan"
Airport Name : "Coldfoot Airport"
Airport Country Code : "US"
Country Name : "United States"
Airport Continent : "NAM"
Continents : "North America"
Departure Date : "6/28/2022"
Arrival Airport : "CXF"
Pilot Name : "Fransisco Hazeldine"
Flight Status : "On Time"

_id: ObjectId('678a2de681bfd3dd7a52b26b')
Passenger ID : "jkXXAX"
First Name : "Elwood"
Last Name : "Catt"
Gender : "Male"
Age : 62
Nationality : "Nicaragua"
Airport Name : "Kugluktuk Airport"
Airport Country Code : "CA"
Country Name : "Canada"
Airport Continent : "NAM"
Continents : "North America"
Departure Date : "12/26/2022"
Arrival Airport : "YCO"
Pilot Name : "Marla Parsonage"
Flight Status : "On Time"

_id: ObjectId('678a2de681bfd3dd7a52b26c')
Passenger ID : "CdUz2g"
             : "Darby"
               "Felgate"
```

Import completed.

Simple Indexes

Go to Indexes and click on Create Index

Firdaus Mulla | L007

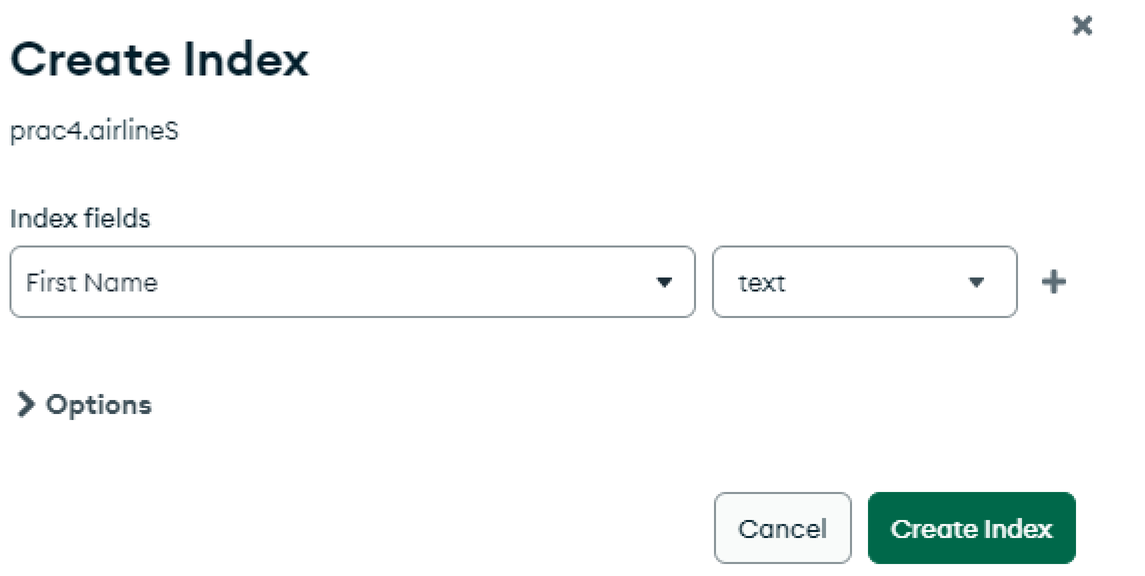


Name and Definition	Type	Size	Usage	Properties
_id_	REGULAR	995.3 KB	2 (since Fri Jan 17 2025)	UNIQUE

Create Index

prac4.airlineS

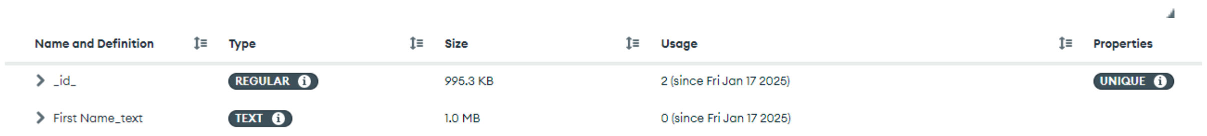
Index fields


Create Index

prac4.airlineS

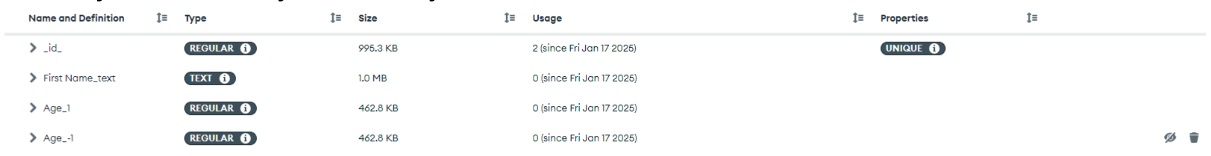
Index fields

First Name | text



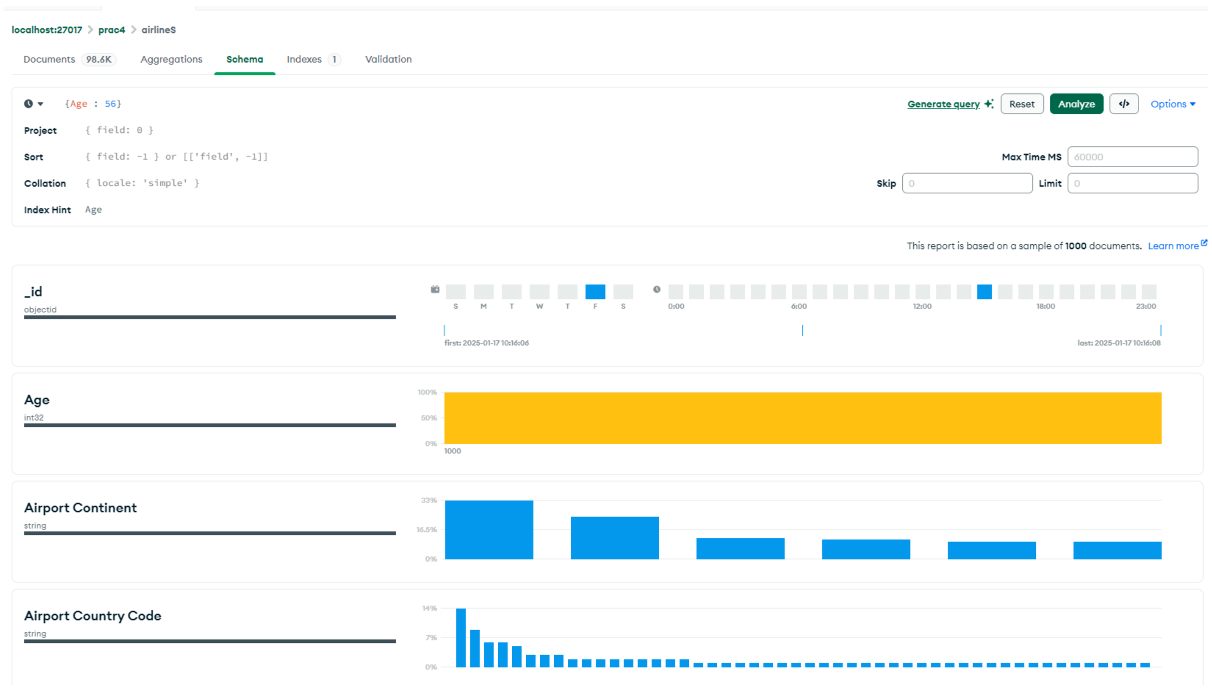
Name and Definition	Type	Size	Usage	Properties
_id_	REGULAR	995.3 KB	2 (since Fri Jan 17 2025)	UNIQUE
First Name_text	TEXT	1.0 MB	0 (since Fri Jan 17 2025)	

Similarly create as many indexes as you want



Name and Definition	Type	Size	Usage	Properties
_id_	REGULAR	995.3 KB	2 (since Fri Jan 17 2025)	UNIQUE
First Name_text	TEXT	1.0 MB	0 (since Fri Jan 17 2025)	
Age_1	REGULAR	462.8 KB	0 (since Fri Jan 17 2025)	
Age_-1	REGULAR	462.8 KB	0 (since Fri Jan 17 2025)	

Firdaus Mulla | L007


localhost:27017 > prac4 > airlineS

Documents 98.6K | Aggregations | Schema | Indexes 1 | Validation

```
{Age : 56}
Project      { field: 0 }
Sort         { field: -1 } or [['field', -1]]
Collation    { locale: 'simple' }
Index Hint   Age
```

Max Time MS 60000
Skip 0 Limit 0

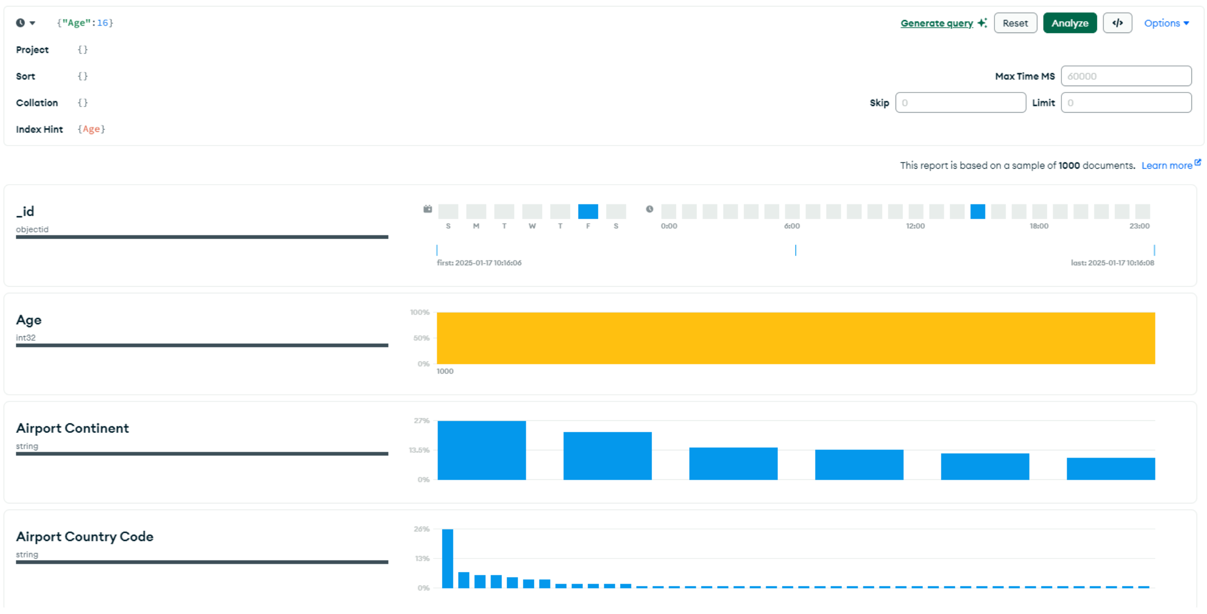
This report is based on a sample of **1000** documents. Learn more

_id
objectid

Age
int32

Airport Continent
string

Airport Country Code
string


```
{"Age":16}
Project      {}
Sort         {}
Collation    {}
Index Hint   {Age}
```

Max Time MS 60000
Skip 0 Limit 0

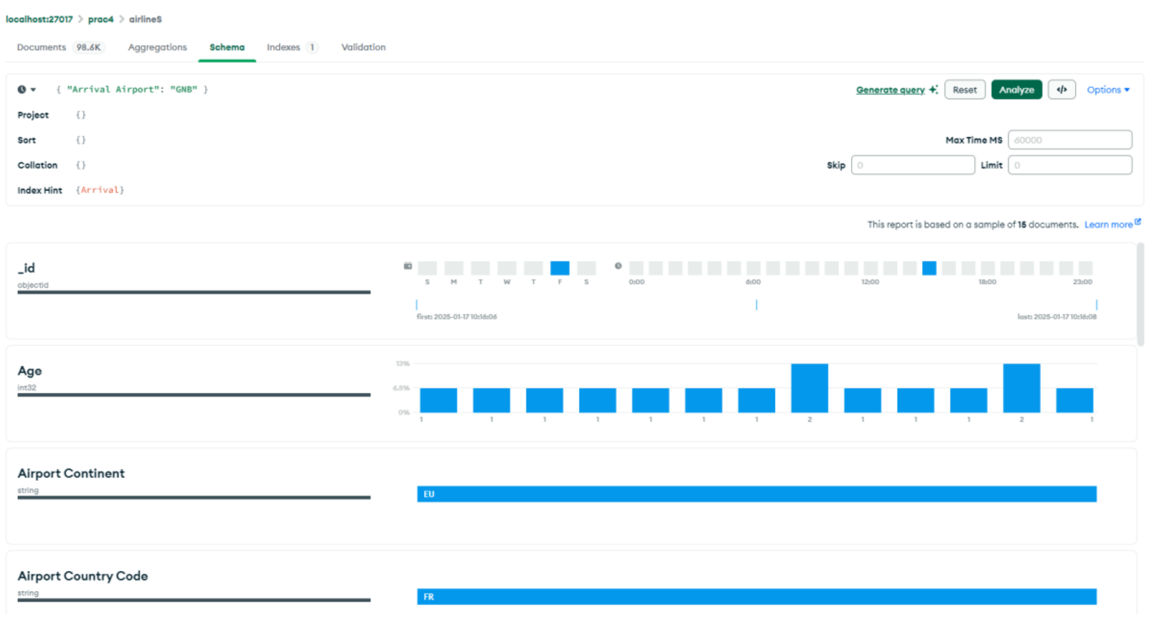
This report is based on a sample of **1000** documents. Learn more

_id
objectid

Age
int32

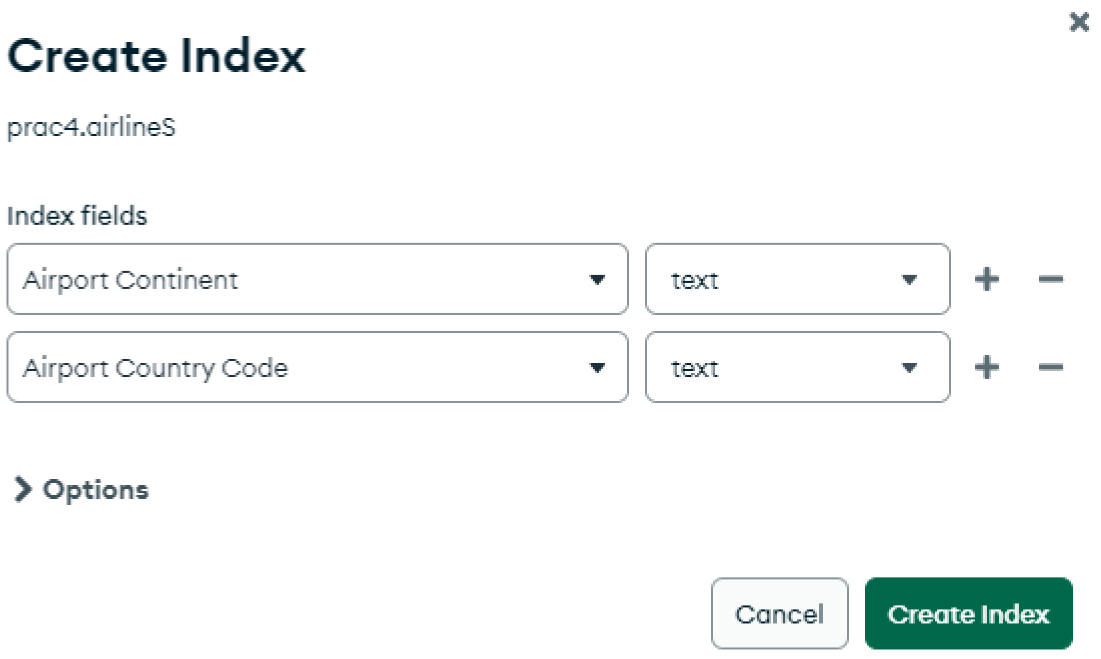
Airport Continent
string

Airport Country Code
string

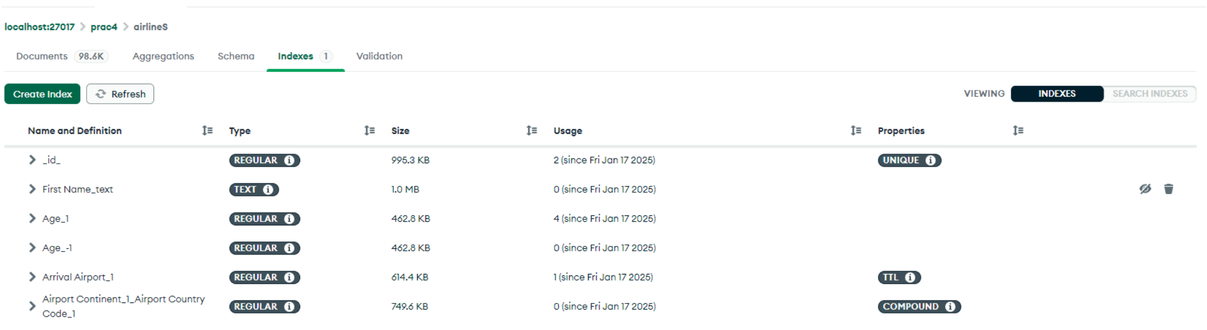


```
{ "Arrival Airport": "GNB" }
Project     {}
Sort        {}
Collation   {}
Index Hint  {Arrival}
Max Time MS 60000
Skip 0   Limit 0
```

Compound Indexes:

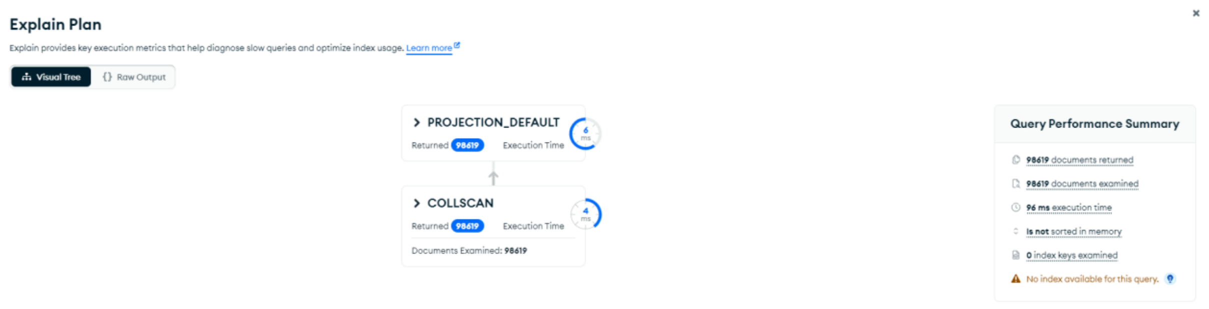


Index fields	
Airport Continent	text
Airport Country Code	text



Name and Definition	Type	Size	Usage	Properties
_id_	REGULAR	995.3 KB	2 (since Fri Jan 17 2025)	UNIQUE
First Name_text	TEXT	1.0 MB	0 (since Fri Jan 17 2025)	
Age_1	REGULAR	462.8 KB	4 (since Fri Jan 17 2025)	
Age_-1	REGULAR	462.8 KB	0 (since Fri Jan 17 2025)	
Arrival Airport_1	REGULAR	614.4 KB	1 (since Fri Jan 17 2025)	TTL
Airport Continent_1_Airport Country Code_1	REGULAR	749.6 KB	0 (since Fri Jan 17 2025)	COMPOUND

Firdaus Mulla | L007

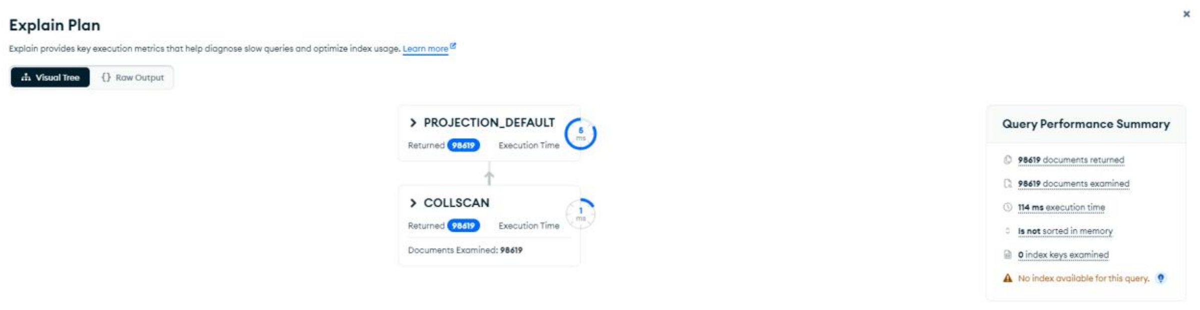

Explain Plan

Explain provides key execution metrics that help diagnose slow queries and optimize index usage. Learn more

PROJECTION_DEFAULT
Returned 98619 Execution Time 6 ms

COLLSCAN
Returned 98619 Execution Time 4 ms
Documents Examined: 98619

Query Performance Summary
98619 documents returned
98619 documents examined
96 ms execution time
Is not sorted in memory
0 index keys examined
No index available for this query.


Explain Plan

Explain provides key execution metrics that help diagnose slow queries and optimize index usage. Learn more

PROJECTION_DEFAULT
Returned 98619 Execution Time 5 ms

COLLSCAN
Returned 98619 Execution Time 1 ms
Documents Examined: 98619

Query Performance Summary
98619 documents returned
98619 documents examined
114 ms execution time
Is not sorted in memory
0 index keys examined
No index available for this query.

****