knowledgehut
upGrad

MongoDB Drivers and Python

Pre-requisites

Hope you have gone through the self-learning content for this session on the PRISM portal.

By the End of this Session:

- Understand the use of drivers for integrating MongoDB with multiple programming languages.
- Install and use PyMongo to work with MongoDB in Python.
- Perform the CRUD operation using PyMongo.
- Create indexes to improve the speed of queries.
- Understand the aggregation framework to create and execute complex pipelines.
- Use GridFS to store large data files in MongoDB.

What’s In It For Me?

Poll Time

Q. Which MongoDB operation is used to remove a single document from a collection based on a specified filter?

a. deleteOne()

b. removeOne()

c. eraseOne()

d. dropOne()

Poll Time

Q. Which MongoDB operation is used to remove a single document from a collection based on a specified filter?

- a. **deleteOne()**
- b. removeOne()
- c. eraseOne()
- d. dropOne()

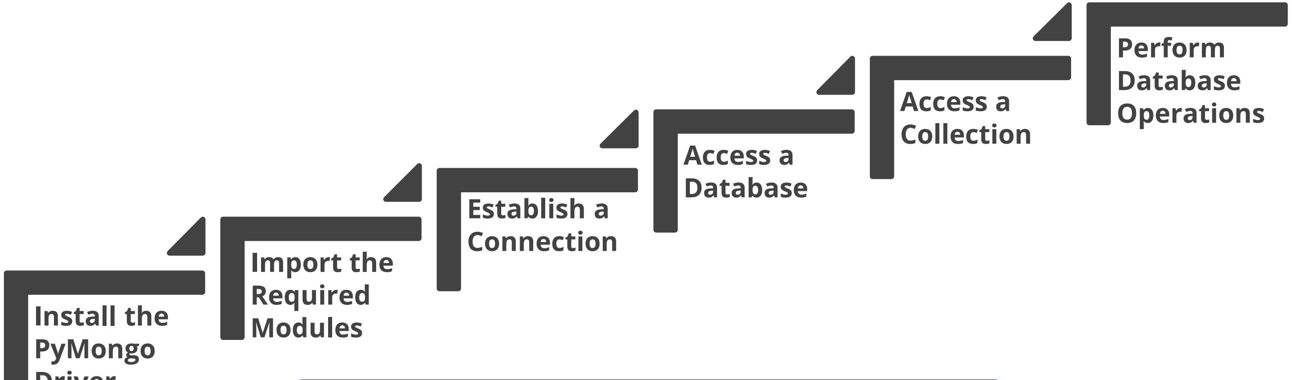
Introduction to MongoDB Drivers

MongoDB drivers are software libraries or APIs (Application Programming Interfaces) that allow developers to interact with MongoDB databases using programming languages.

Popular Driver Support

- Python
- Java
- Node.js (JavaScript)
- C#
- Ruby
- PHP

Using Python to Interface with MongoDB



Install the PyMongo Driver

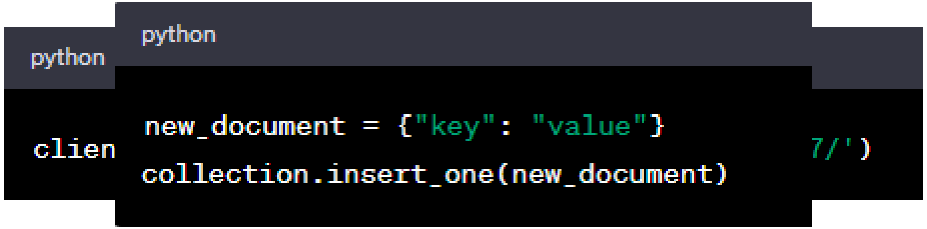
Import the Required Modules

Establish a Connection

Access a Database

Access a Collection

Perform Database Operations



python

```
clien                                                  7/')
```

python

```python
new_document = {"key": "value"}
collection.insert_one(new_document)
```

Pop Quiz

Q. What is PyMongo?

a. A programming language

b. A query language for MongoDB

c. The official MongoDB driver for Python

d. A database management system

Pop Quiz

Q. What is PyMongo?

a. A programming language

b. A query language for MongoDB

c. The official MongoDB driver for Python

d. A database management system

Demo - Connecting to MongoDB

Poll Time

Q. What is the purpose of the **Cursor** returned by **find()** in PyMongo?

a. It represents a collection schema

b. It allows you to execute database commands

c. It helps iterate over query results

d. It handles authentication and authorization

Poll Time

Q. What is the purpose of the **Cursor** returned by **find()** in PyMongo?

a. It represents a collection schema

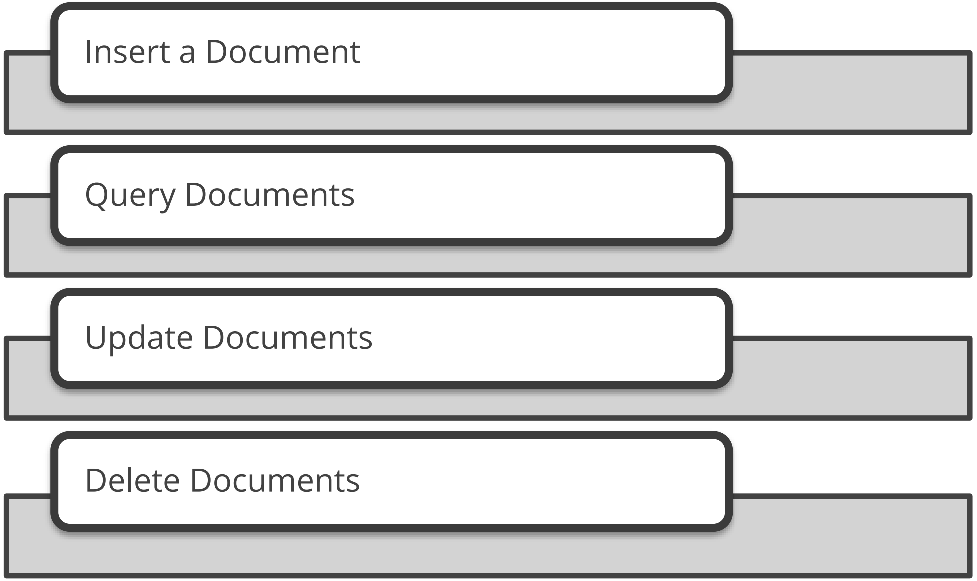
b. It allows you to execute database commands

c. **It helps iterate over query results**

d. It handles authentication and authorization

Q&A

Performing Simple Operations with Python



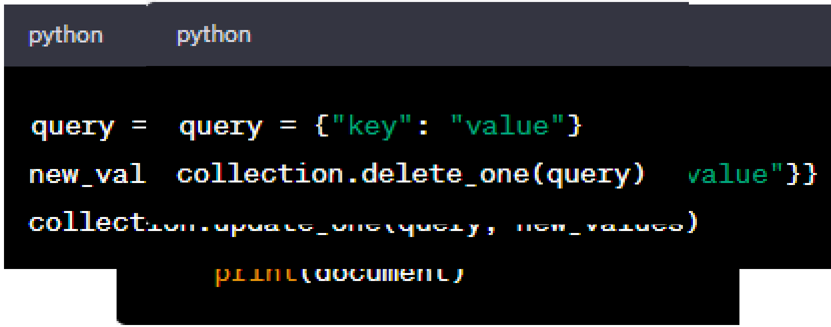
Insert a Document

Query Documents

Update Documents

Delete Documents

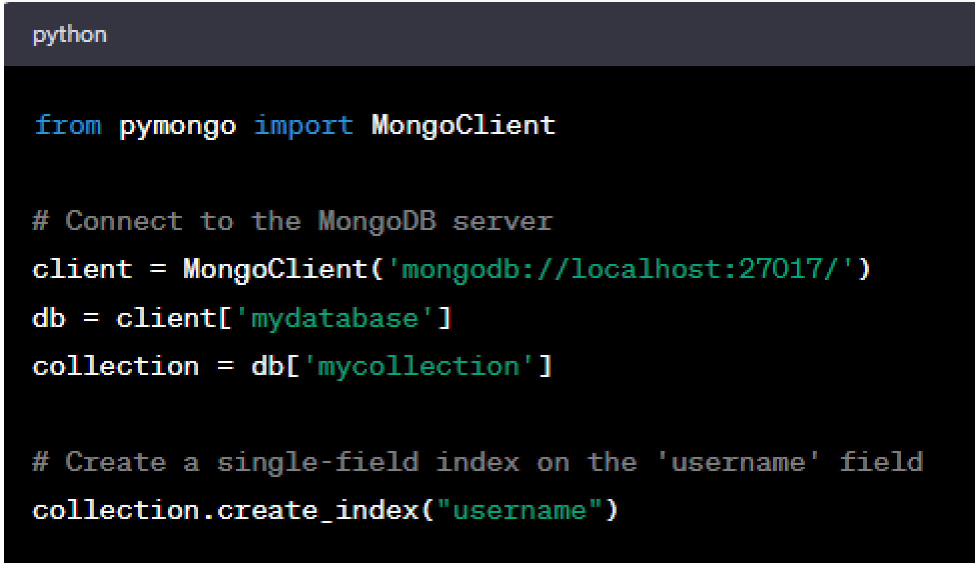
Example



```
query =  query = {"key": "value"}
new_val  collection.delete_one(query)  value"}}
collection.update_one(query, new_values)
        print(document)
```

Indexing in MongoDB

Indexing in MongoDB is a database optimization technique that improves query performance by creating data structures that allow the database to quickly locate and retrieve specific documents based on certain fields.



```python
from pymongo import MongoClient

# Connect to the MongoDB server
client = MongoClient('mongodb://localhost:27017/')
db = client['mydatabase']
collection = db['mycollection']

# Create a single-field index on the 'username' field
collection.create_index("username")
```

Pop Quiz

Q. Which of the following is used to insert a document into a **MongoDB** collection using **PyMongo**?

a. collection.insert()

b. collection.create_document()

c. collection.add()

d. collection.insert_one()

Pop Quiz

Q. Which of the following is used to insert a document into a **MongoDB** collection using **PyMongo**?

a. collection.insert()

b. collection.create_document()

c. collection.add()

[✓ **d. collection.insert_one()**]

Aggregation Framework

The Aggregation Framework in MongoDB is a powerful tool for performing data processing and transformation operations on your data within the database.

$match

- Filters documents based on specified conditions.

$group

- Groups documents by a specified key and calculates aggregate values for each group.

$sort

- Sorts documents.

$project

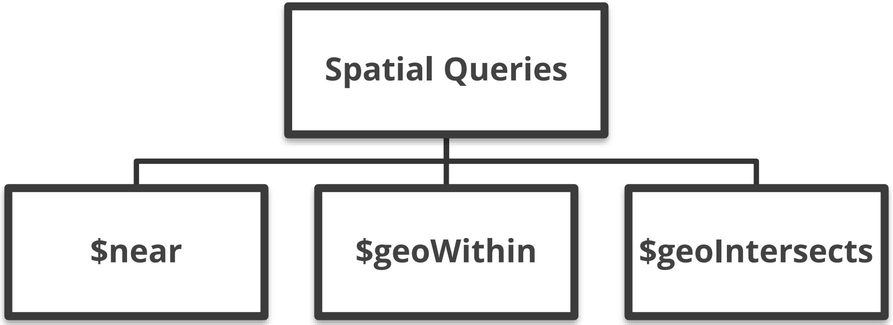
- Shapes the output by specifying which fields to include or exclude.

$limit

- Limit the number of documents

Geospatial Data Analysis

Geospatial data analysis involves working with data that has a geographic or spatial component.

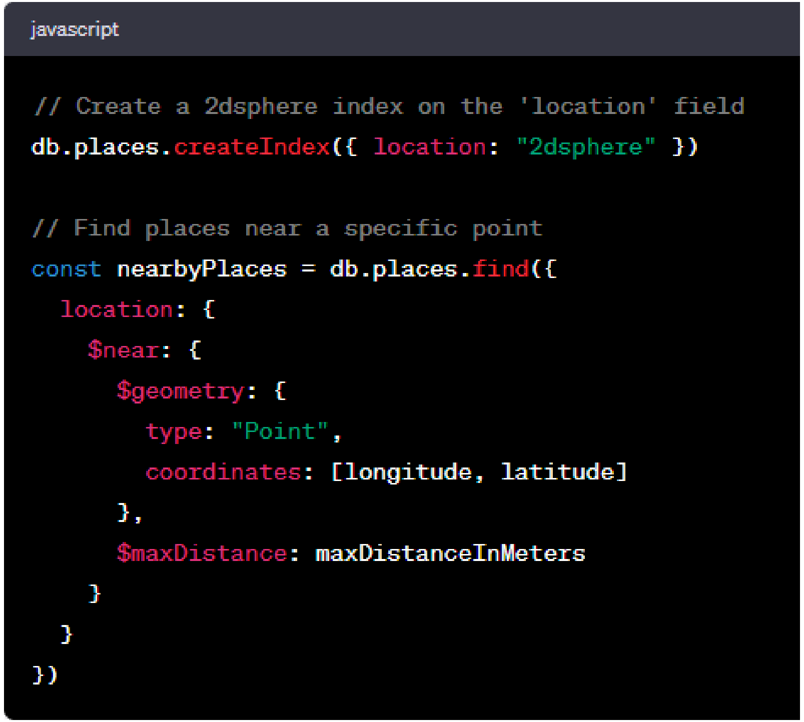


Spatial Queries

$near

$geoWithin

$geoIntersects

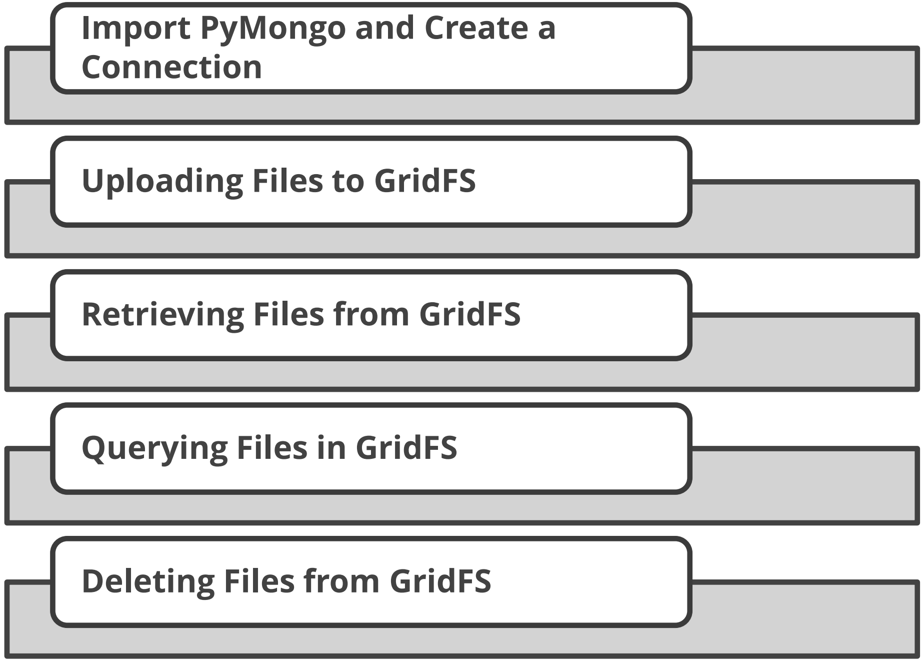


```javascript
// Create a 2dsphere index on the 'location' field
db.places.createIndex({ location: "2dsphere" })

// Find places near a specific point
const nearbyPlaces = db.places.find({
  location: {
    $near: {
      $geometry: {
        type: "Point",
        coordinates: [longitude, latitude]
      },
      $maxDistance: maxDistanceInMeters
    }
  }
})
```

Working with GridFS

- The BSON format has a document size limit of 16 MB.
- GridFS is a specification for storing and retrieving large files (such as images, videos, audio files, and documents) in MongoDB, which surpasses the above size limit.



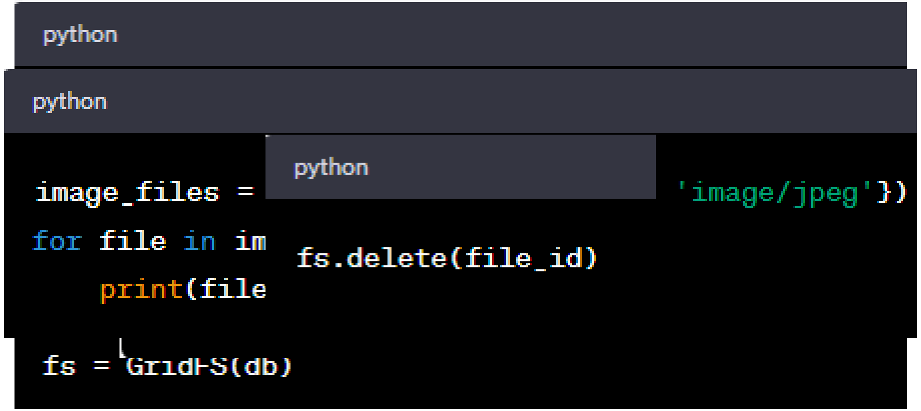
Import PyMongo and Create a Connection

Uploading Files to GridFS

Retrieving Files from GridFS

Querying Files in GridFS

Deleting Files from GridFS



python

```python
fs = GridFS(db)
```

python

```python
image_files =                                    'image/jpeg'})
for file in im
    print(file
```

python

```python
fs.delete(file_id)
```

Poll Time

Q. What type of data can be stored using GridFS in PyMongo?

a. Only small text-based data

b. Large files that exceed the BSON size limit

c. Only images and videos

d. Small JSON documents

Poll Time

Q. What type of data can be stored using GridFS in PyMongo?

a. Only small text-based data

b. Large files that exceed the BSON size limit

c. Only images and videos

d. Small JSON documents

Q&A

Activity 1

Pre-requisites:

- Install the **PyMongo** package using **pip install PyMongo**.
- Have a MongoDB server running locally or remotely.

Scenario:

Create a basic command-line to-do list application using **PyMongo** where users can add, list, and mark tasks as done.

Setup:

- Import the necessary modules: PyMongo, datetime.
- Create a MongoDB client and connect to your database.

Create Functions:

Write functions to perform the following actions:

- Add a task with a description and due date to the collection
- List all tasks
- Mark a task as done
- Delete a task

Activity 1

User Interaction:

- Implement a basic command-line interface that allows users to interact with the application.
- Display a menu with options for adding, listing, marking, and deleting tasks.

Implement the Application:

- Display tasks in a readable format, showing their descriptions, due dates, and whether they are done or not.

Error Handling:

- Handle cases where users input invalid options or tasks.

Summary

- ✓ Drivers facilitate interaction between the application and **MongoDB server**.
- ✓ **PyMongo** is the Python driver for interacting with MongoDB.
- ✓ We can perform all **CRUD** operations using **PyMongo**.
- ✓ Indexing in MongoDB is a way to improve query performance.
- ✓ The Aggregation Framework is used to perform complex data processing and analysis tasks directly within the database.
- ✓ GridFS is a method for storing and retrieving large files.

Next Session:
Case Study on MongoDB

THANK YOU!

Please complete your assessments and review the self-learning content for this session on the **PRISM** portal.

knowledgehut
upGrad

Case Study on MongoDB

Pre-requisites

Hope you have gone through the self-learning content for this session on the PRISM portal.

By the End of this Session:

- Create a MongoDB database which contains customer and accounts collection.
- Perform CRUD operations on these collections.
- Write and execute queries to analyze the underlying data.
- Save the results to output files.

Recap

Poll Time

Q. Which aggregation stage is used to filter documents based on specified conditions?

a. $sort

b. $group

c. $match

d. $project

Poll Time

Q. Which aggregation stage is used to filter documents based on specified conditions?

a. $sort

b. $group

[✓ **c. $match**]

d. $project

Case Study

Introduction to the Case Study

Create a MongoDB database to store and analyze customer and their respective account information.

Problem Statement

Tasks to Perform

- Create collections to store customer and account information
- Perform CRUD operations on the above collections.
- Write queries to analyze the underlying data
- Save the results to files for further reporting.

Pop Quiz

Q. Which pymongo method is used to store a file in GridFS?

a. collection.insert()
b. fs.put()
c. collection.save()
d. fs.create_file()

Pop Quiz

Q. Which pymongo method is used to store a file in GridFS?

a. collection.insert()

[✓ **b. fs.put()**]

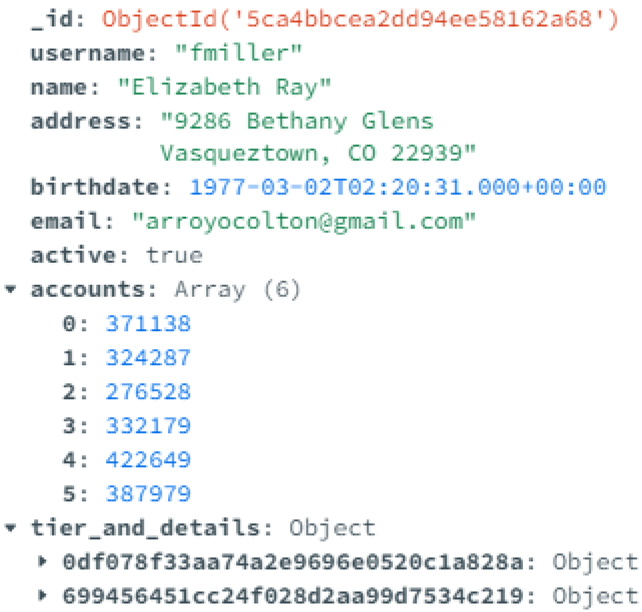
c. collection.save()

d. fs.create_file()

Understanding the Data

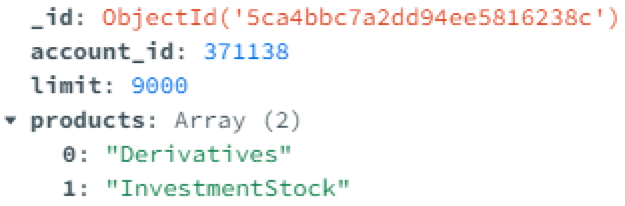
Sneak Peak into the Data

Customer Collection



```
  _id: ObjectId('5ca4bbcea2dd94ee58162a68')
  username: "fmiller"
  name: "Elizabeth Ray"
  address: "9286 Bethany Glens
           Vasqueztown, CO 22939"
  birthdate: 1977-03-02T02:20:31.000+00:00
  email: "arroyocolton@gmail.com"
  active: true
▾ accounts: Array (6)
    0: 371138
    1: 324287
    2: 276528
    3: 332179
    4: 422649
    5: 387979
▾ tier_and_details: Object
  ▸ 0df078f33aa74a2e9696e0520c1a828a: Object
  ▸ 699456451cc24f028d2aa99d7534c219: Object
```

Account Collection



```
  _id: ObjectId('5ca4bbc7a2dd94ee5816238c')
  account_id: 371138
  limit: 9000
▾ products: Array (2)
    0: "Derivatives"
    1: "InvestmentStock"
```

Poll Time

Q. Which of the following is NOT a valid aggregation operation in MongoDB?

a. $match

b. $query

c. $group

d. $project

Poll Time

Q. Which of the following is NOT a valid aggregation operation in MongoDB?

a. $match

[✓ **b. $query**]

c. $group

d. $project

Q&A

Hands-on: Case Study Questions

Pop Quiz

Q. Which of the following features is NOT provided by MongoDB Compass?

a. Visual query builder

b. Aggregation pipeline builder

c. Automatic code generation for applications

d. Schema validation and analysis

Pop Quiz

Q. Which of the following features is NOT provided by MongoDB Compass?

a. Visual query builder

b. Aggregation pipeline builder

[✓ **c. Automatic code generation for applications**]

d. Schema validation and analysis

Q&A

Activity 1

Pre-requisites:

- Install the **PyMongo** package using **pip install PyMongo**.
- Have a MongoDB server running locally or remotely.

Scenario:

Use the customer and account collections and run the below queries on the data.

- Write a query to all the customers that are active. Display only the name and the email of the customer and store the result in csv file.
- Write a query to retrieve all customers located in the New Mexico State and store the result in csv file.
- Write a query to find the customer with the lowest total account limit and store the result in csv file.

Summary

- ✓ MongoDB Compass can be used to directly insert data from JSON files.
- ✓ You can execute commands using mongodb shell or you can use the GUI in the MongoDB compass.
- ✓ CRUD operation allow us to manipulate the underlying data.
- ✓ We can output the results to csv files.

Session Feedback

Next Session:
Introduction to Tableau

THANK YOU!

Please complete your assessments and review the self-learning content for this session on the **PRISM** portal.

knowledgehut
upGrad