

# MongoDB and Python Integration

Applied Database Technologies DS-532  
Ameya Dalvi  
Prof. Olga Scrivner

# MongoClient

1. The MongoClient() class for the PyMongo driver library for MongoDB creates client instances. Its main function is to enable you to connect to the MongoDB database efficiently and effortlessly. Whether you're changing data or retrieving it, every connection matters.
1. **from pymongo import MongoClient**

# Workshop Setup

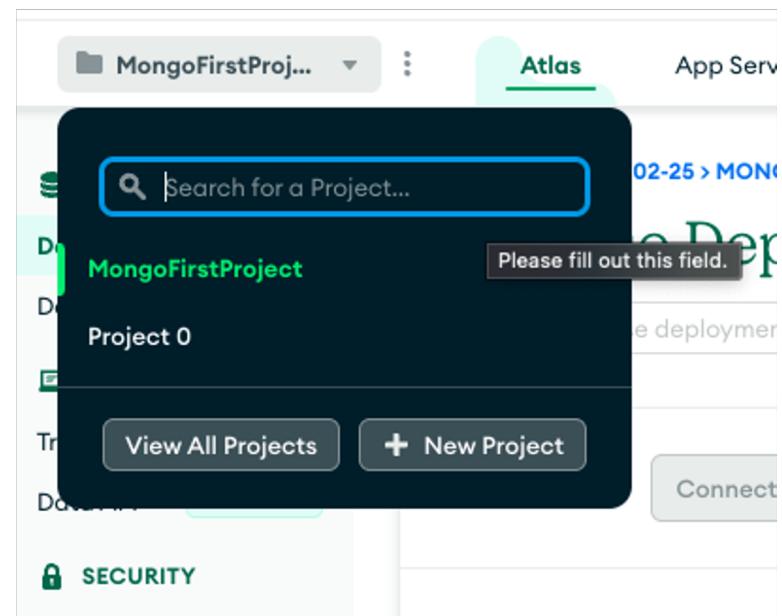
## 1. Create an Atlas account:

<https://www.mongodb.com/cloud/atlas>

Note - you can use the free tier for this workshop.

## 1. Create a project:

Create a project, give it a name (no specific restrictions here, but choose something descriptive like "mongo-python-workshop")



Follow along these steps:

## Step1: Name your project

The screenshot shows the 'Create a Project' step 1 interface. At the top, there's a navigation bar with 'Ameya's Org...', 'Access Manager', 'Billing', 'All Clusters', 'Get Help', and a user dropdown. On the left, a sidebar lists 'Projects' (selected), 'Alerts', 'Activity Feed', 'Settings', 'Integrations', 'Access Manager', 'Billing', 'Support', and 'Live Migration'. The main area has a title 'AMEYA'S ORG - 2022-02-25 > PROJECTS' and 'Create a Project'. It features two buttons: 'Name Your Project' (highlighted) and 'Add Members'. Below is a section titled 'Name Your Project' with the sub-instruction 'Project names have to be unique within the organization (and other restrictions)'. A text input field contains 'mongo-python-workshop'. At the bottom are 'Cancel' and 'Next' buttons.

## Step2: Add members, Set permissions

The screenshot shows the 'Create a Project' step 2 interface. The layout is similar to the first step. The sidebar and top navigation are identical. The main area has a title 'AMEYA'S ORG - 2022-02-25 > PROJECTS' and 'Create a Project'. It shows a green checkmark next to 'Name Your Project' and the 'Add Members' button. Below is a section titled 'Add Members and Set Permissions' with the sub-instruction 'Invite new or existing users via email address...'. A text input field contains 'abdalvi@iu.edu (you)'. To its right is a dropdown menu set to 'Project Owner'. At the bottom are 'Cancel', 'Go Back', and 'Create Project' buttons.

### 3. Set up a cluster

After creating a project, set up a **Cluster**

Make sure you name it! You may need to scroll to the bottom of the page. This is a one-time setup: once your cluster is created, you won't be able to change its name.

Click on Build a Database

The screenshot shows the MongoDB Atlas interface. At the top, there are navigation links: 'Ameya's Org...', 'Access Manager', 'Billing', 'All Clusters', 'Get Help', and 'Ameya'. Below this is a search bar with 'mongo-python...' and a dropdown menu showing 'Atlas', 'App Services', and 'Charts'. On the left, a sidebar has sections for 'DEPLOYMENT' (selected), 'Database' (highlighted in green), 'Data Lake', 'DATA SERVICES' (Triggers, Data API), and 'SECURITY' (Database Access, Network Access, Advanced). A 'PREVIEW' button is also present. The main content area is titled 'Database Deployments' and features a large 'Create a database' button with a plus sign icon. Below it, the text 'Choose your cloud provider, region, and specs.' and a 'Build a Database' button. At the bottom, a note says 'Once your database is up and running, live migrate an existing MongoDB database into Atlas with our [Live Migration Service](#)'.

Click on the Create button for the Shared db

The screenshot shows the 'Deploy a cloud database' page. It features three deployment options: 'Serverless', 'Dedicated', and 'Shared'. Each option has a 'Create' button and a price starting point. A dashed arrow points from the 'Shared' option towards the bottom right, with the text 'I'll do this later' written below it. At the bottom right, there is a link 'Advanced Configuration Options'.

Deployment Type	Starting at
Serverless	\$0.30/1M reads
Dedicated	\$0.08/hr*
Shared	FREE



Ameya ▾

CLUSTERS &gt; CREATE A SHARED CLUSTER

## Create a Shared Cluster

Welcome to MongoDB Atlas! We've recommended some of our most popular options to your needs. For more information, check our [documentation](#).

Use the Shared cluster option

PREVIEW Serverless

Dedicated

FREE Shared

For learning and exploring MongoDB in a sandbox environment. Basic configuration controls.

No credit card required to start. Upgrade to dedicated clusters for full functionality.  
Explore with sample datasets. Limit of one free cluster per project.

Use the AWS Cloud provider



★ Recommended region ⓘ ⓘ Dedicated tier region ⓘ

Use the N.Virginia (us-east-1) region



EUROPE

Frankfurt (eu-central-1) ★

N. Virginia (us-east-1) ★

Ohio (us-east-2) ★ ⓘ

Stockholm (eu-north-1) ★

Paris (eu-west-3) ★

AUSTRALIA

Sydney (ap-southeast-2) ★

ASIA

Tokyo (ap-northeast-1)

FREE

Your M0 cluster is ideal for experimenting in a limited sandbox. You can upgrade to a production cluster anytime.

Back

Create Cluster



## Create a cluster name:

Cluster Name

One time only: once your cluster is created, you won't be able to change its name.

Cluster

Cluster names can only contain ASCII letters, numbers, and hyphens.

## 4. Create user and provide r/w access

Create a new database user name:

**pymongo\_workshop\_read\_write**

You can autogenerate a password or choose one (if you autogenerate it, save it somewhere) give this user write privileges, but not full admin

**Password Authentication**

SHOW

Autogenerate Secure Password Copy

**Database User Privileges**

Configure role based access control by assigning database users a mix of one built-in role, multiple custom roles, and multiple specific privileges. A user will gain access to all actions within the roles assigned to them, not just the actions those roles share in common. **You must choose at least one role or privilege.** [Learn more about roles.](#)

**Built-in Role** 1 SELECTED ▲

Select one [built-in role](#) for this user.

**Custom Roles** ▼

Select your [pre-defined custom role\(s\)](#). Create a custom role in the [Custom Roles](#) tab.

**Specific Privileges** ▼

Select multiple privileges and what database and collection they are associated with. Leaving collection blank will grant this role for all collections in the database.

**Restrict Access to Specific Clusters/Data Lakes**

Enable to specify the resources this user can access. By default, all resources in this project are accessible.

**Temporary User**

This user is temporary and will be deleted after your specified duration of 6 hours, 1 day, or 1 week.

Cancel Add User

## 5. Provide network access

You can either make this world-readable or limit to specific IP address.

To limit to your current connection, use your IP address (note - if you don't have a persistent or permanent IP address, you won't have access next time you log in. If this is the case, you'll need to update your network access next time you try to read from this account outside your Atlas login).

To make it world-readable (no IP restrictions) 0.0.0.0/0

Add entries to your IP Access List

Only an IP address you add to your Access List will be able to connect to your project's clusters. You can manage existing IP entries via the [Network Access Page](#).

IP Address	Description	
<input type="text" value="Enter IP Address"/>	<input type="text" value="Enter description"/>	<button>Add Entry</button>
		<button>Add My Current IP Address</button>

---

IP Access List	Description	
66.244.81.31/32	My IP Address	REMOVE

**After configuring your cluster details, you will be redirected to this page :**

The screenshot shows the MongoDB Atlas interface. On the left, a sidebar menu includes 'DEPLOYMENT', 'Database' (selected), 'Data Lake', 'DATA SERVICES', 'Triggers', 'Data API' (selected), 'PREVIEW' (highlighted in green), 'SECURITY', 'Database Access', 'Network Access', 'Advanced', and 'New On Atlas'. The main content area is titled 'AMEYA'S ORG - 2022-02-25 > MONGO-PYTHON-WORKSHOP' and 'Database Deployments'. It features a search bar 'Find a database deployment...' and a 'Create' button. Below this, a cluster summary for 'Cluster1' is displayed, showing metrics like R/W connections, data size, and network activity. At the bottom, detailed cluster information is provided:

VERSION	REGION	CLUSTER TIER	TYPE	BACKUPS	LINKED APP SERVICES
5.0.8	AWS / N. Virginia (us-east-1)	M0 Sandbox (General)	Replica Set - 3 nodes	Inactive	None Linked

**Click on Connect, to get the MongoDB connection string:In your cluster**

Select "connect your application"  
choose "python, 3.6 or later" and copy it  
your connection string should look like this (the URI should be slightly different): **Store the connection string**

`mongodb+srv://pymongo_workshop_read_write:<password>@cluster1.ndqga.mongodb.net/?retryWrites=true&w=majority`

### Connect to Cluster1

✓ Setup connection security   Choose a connection method   Connect

**Choose a connection method** [View documentation](#)

Get your pre-formatted connection string by selecting your tool below.

**Connect with the MongoDB Shell**  
Interact with your cluster using MongoDB's interactive Javascript interface

**Connect your application**  
Connect your application to your cluster using MongoDB's native drivers

**Connect using MongoDB Compass**  
Explore, modify, and visualize your data with MongoDB's GUI

[Go Back](#) [Close](#)

### Connect to Cluster1

✓ Setup connection security   ✓ Choose a connection method   Connect

**1 Select your driver and version**

DRIVER	VERSION
Node.js	4.1 or later

**2 Add your connection string into your application code**

Include full driver code example

```
mongodb+srv://pymongo_workshop_read_write:<password>@cluster1.ndqga.mongodb.net/?retryWrites=true&w=majority
```

Replace `<password>` with the password for the `pymongo_workshop_read_write` user. Ensure any option params are [URL encoded](#).

Having trouble connecting? [View our troubleshooting documentation](#)

[Go Back](#) [Close](#)

## 6. Load COVID-19 data into MongoDB on Atlas

We'll use the coronavirus research dataset from the Semantic Scholar team at the Allen Institute.

Go to:

[https://ai2-semanticscholar-cord-19.s3-us-west-2.amazonaws.com/historical\\_releases.html](https://ai2-semanticscholar-cord-19.s3-us-west-2.amazonaws.com/historical_releases.html)

For the workshop, we'll only upload a subset of this data.

We are looking for something small, so we will download the original release, **cord-19\_2020-03-13.tar.gz (0.3 GB)**

2020-03-13	<b>cord-19_2020-03-13.tar.gz</b> (0.3GB)	a36fe181	8fbea927
------------	--	----------	----------

Untar it, you should get a folder named:

**2020-13-13**, within this folder, expand the **noncomm-use-subset**

## 7. Add the Data to a Collection

Back in your Atlas cluster:

Click on **Browse Collections**

Click on **Collections** tab

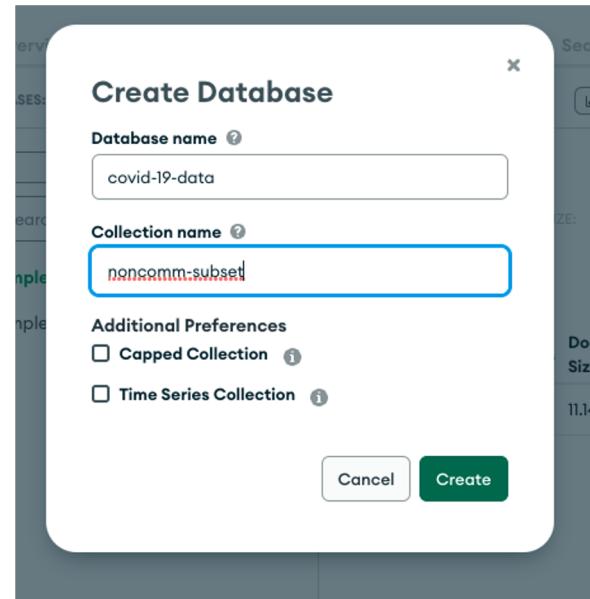
Click on the **+ Create Database** tab to create a database

db name: covid19-data collection name: noncomm-subset

This will basically be an empty database with 0 records currently.

We will insert the downloaded CORD-19 data to it using the

jupyter notebook



AMEYA'S ORG - 2022-02-25 > MONGO-PYTHON-WORKSHOP > DATABASES

**Cluster1**

VERSION 5.0.8 REGION AWS N. Virginia (us-east-1)

Overview Real Time Metrics **Collections** Search Profiler Performance Advisor Online Archive Cmd Line Tools

DATABASES: 11 COLLECTIONS: 24

[+ Create Database](#) [VISUALIZE YOUR DATA](#) [REFRESH](#)

Search Namespaces

- covid-19
- covid-19-data**
  - noncomm-subset**
  - sample\_airbnb
  - sample\_analytics
- sample\_analytics

**covid-19-data.noncomm-subset**

STORAGE SIZE: 4KB TOTAL DOCUMENTS: 0 INDEXES TOTAL SIZE: 4KB

[Find](#) [Indexes](#) [Schema Anti-Patterns](#) [Aggregation](#) [Search Indexes](#)

FILTER { Field: 'value' } [OPTIONS](#) [Apply](#) [Reset](#)

QUERY RESULTS: 0 [INSERT DOCUMENT](#)

Switching to the notebook...