



Big Data, Techniques, and Platforms

Using MongoDB and Studio3T

The purpose of this tutorial is to guide you to use MongoDB and Studio3T.

Notice that you must download and install :

- **MongoDB**
- **Studio3T**

Studio 3T¹ will be used as a client to connect to a MongoDB server instance. Notice that the trial license will expire within 30 days but you can ask for an academic license at this link using your student address :

Studio3T request for academic license.

Download the files and store them in a directory of your computer. I suggest you to create a directory called **data**.

1 INSTALL MONGODB SERVER

First, you need to install MongoDB Community Edition. The instructions are available at :

- Linux Install
- MacOS Install
- Windows Install.

2 HOW TO START WITH MONGODB SERVER

Launch an instance of MongoDB server, following the procedure needed by your machine : Linux, MacOS, and Windows.

1. <https://studio3t.com/features/>

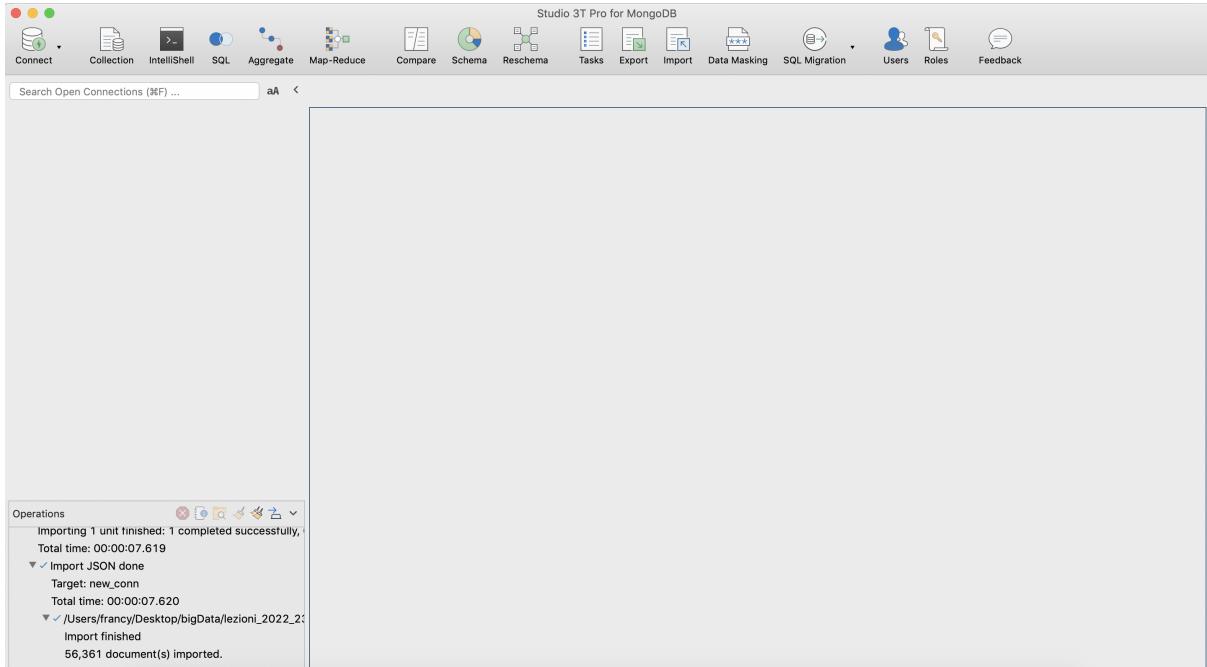
3 INTERACT WITH MONGODB USING STUDIO3T

To query MongoDB database, it is possible to use a client. We use Studio 3T's through IntelliShell. This is a client that makes easier to interact with MongoDB, to show in a better format the result of the query, to inspect the database.

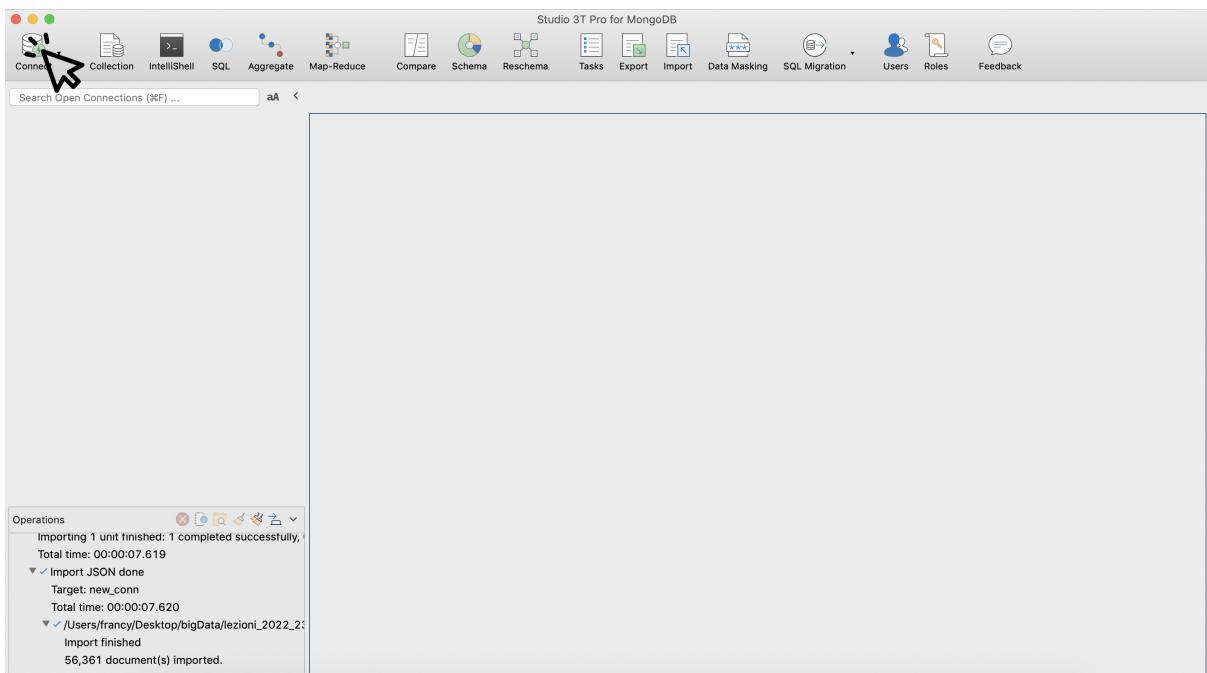
3.1 CONNECT TO THE DATABASE

- Launch Studio 3T
- Click on "Connect"
- Click on "New Connection"
- Click on "Next" without any change
- Click on "Save" without any change
- Click on "Connect" without any change
- You will see the Databases appearing on the left side of the screen

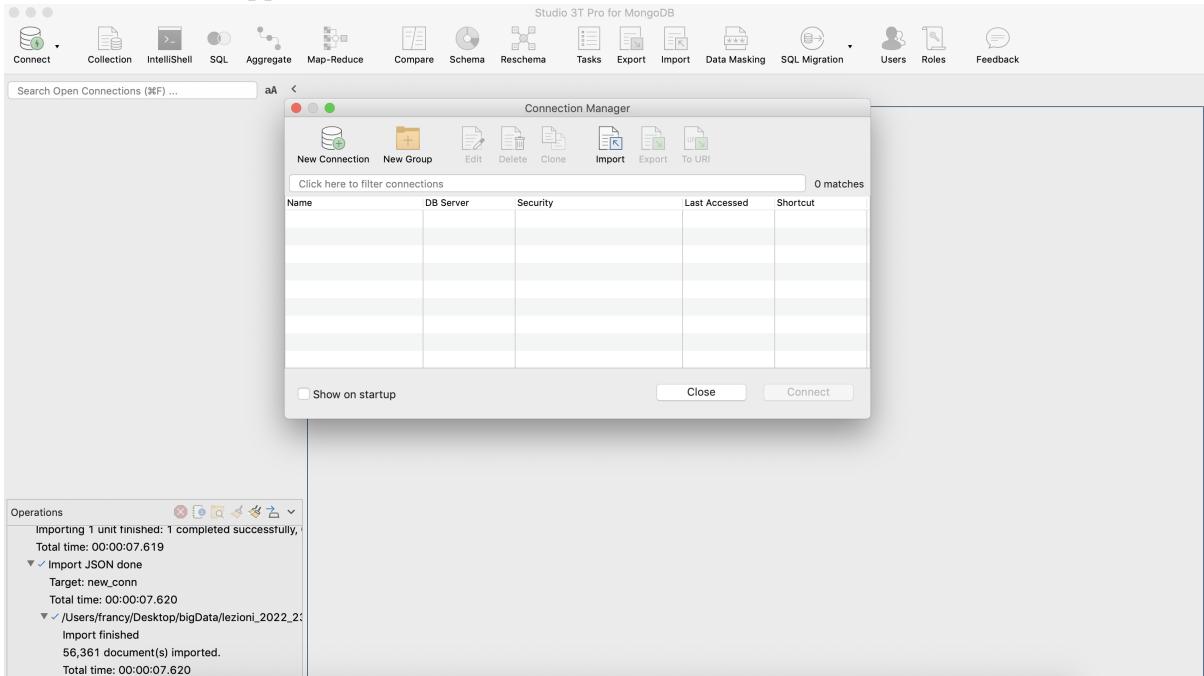
Open Studio3T :



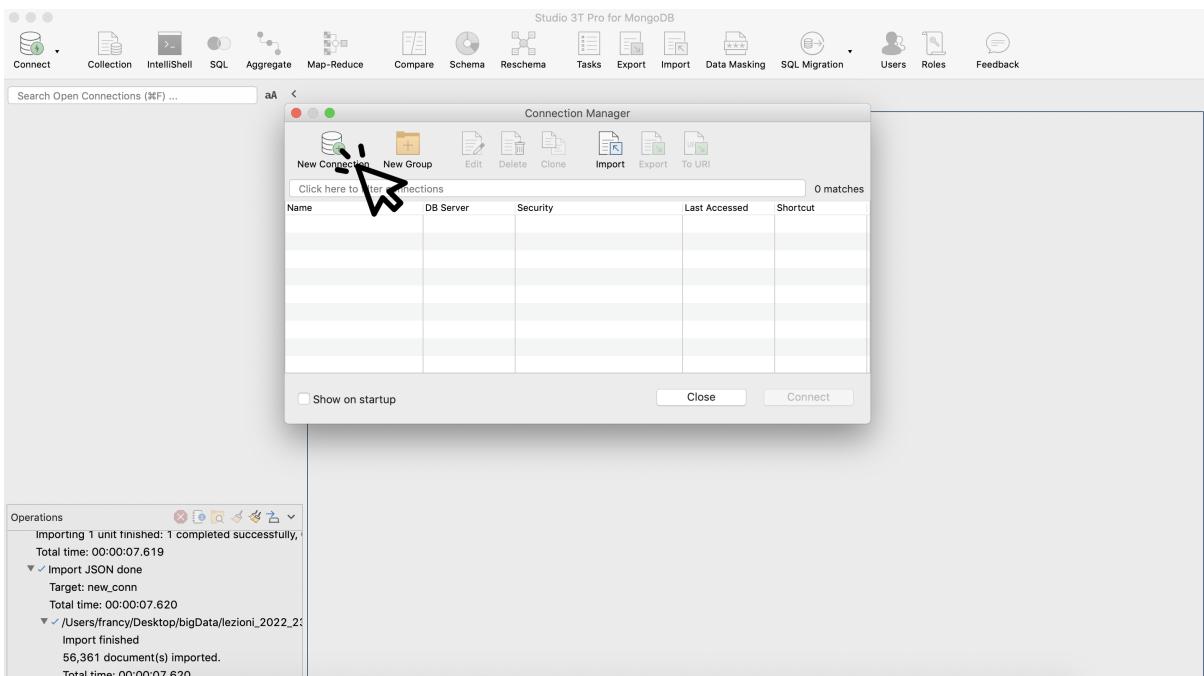
Click on "Connect" :



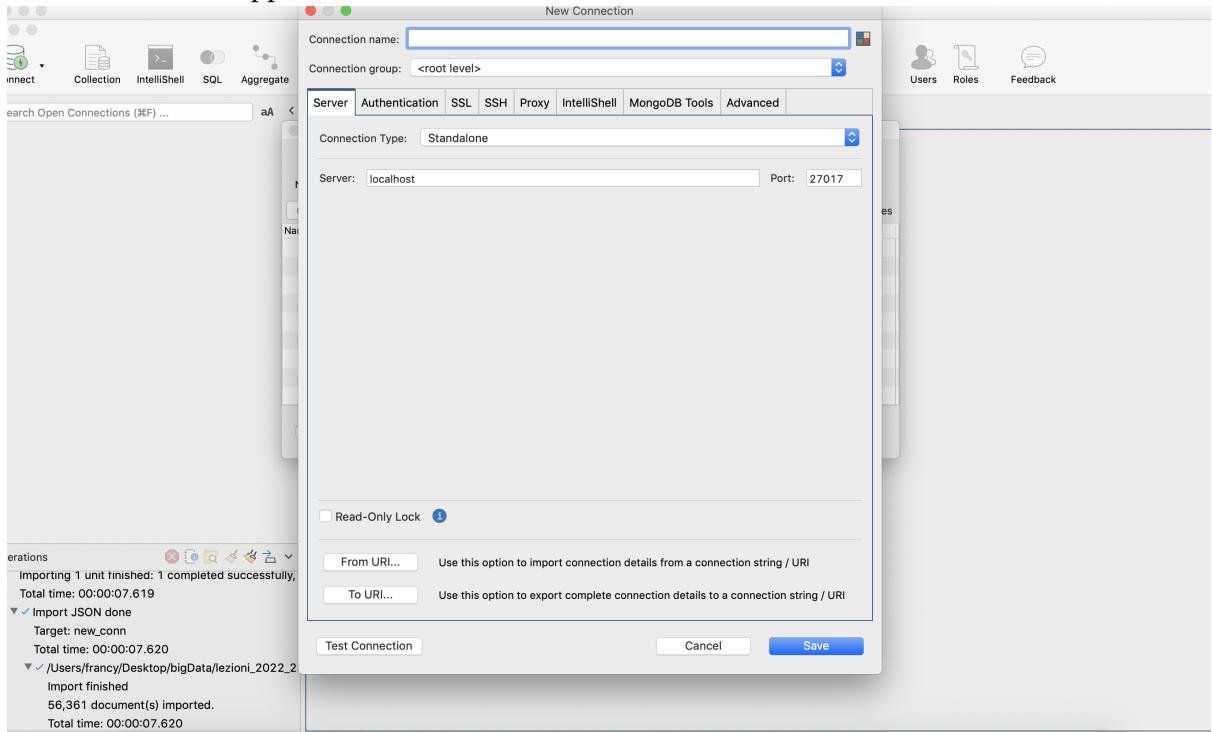
This new form will appear :



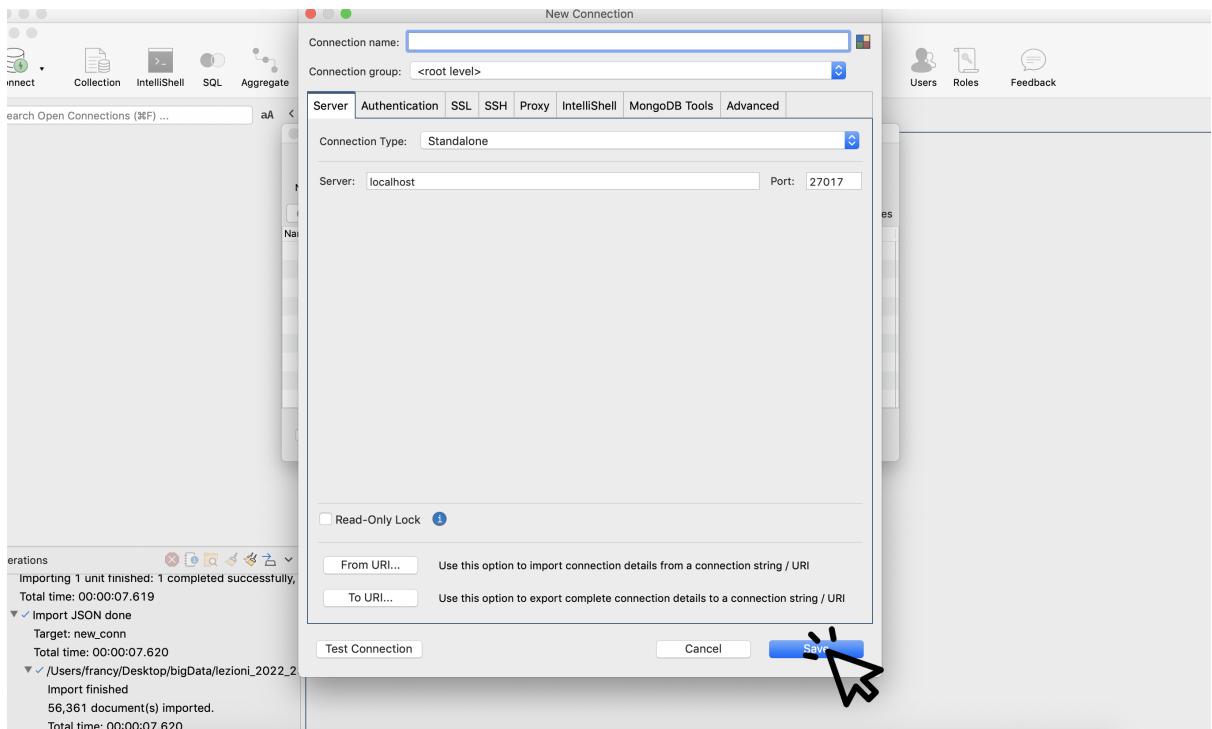
Click on "New Connection" :



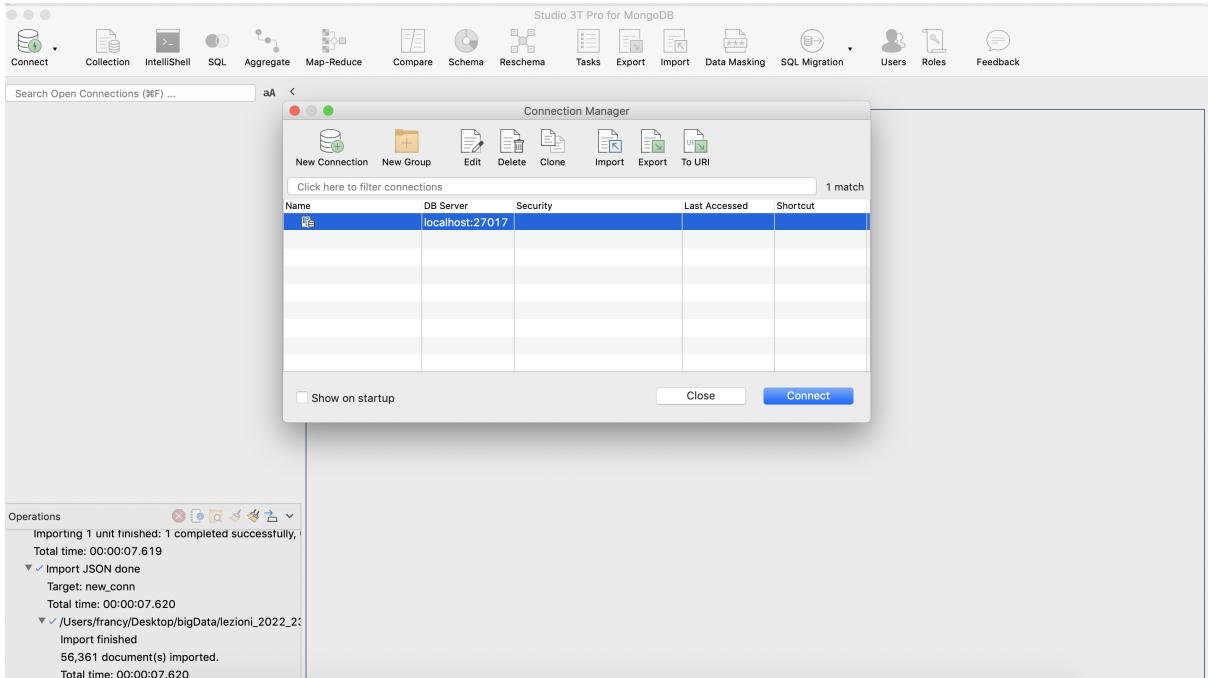
This new form will appear :



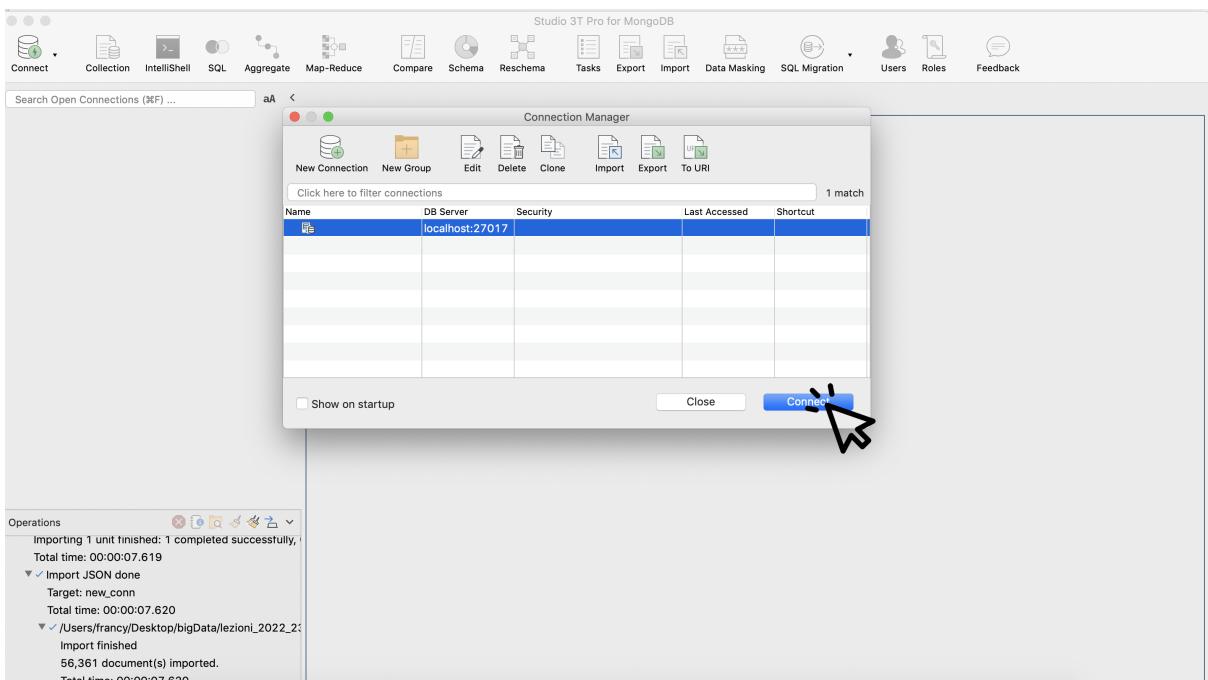
Click on "Next" :



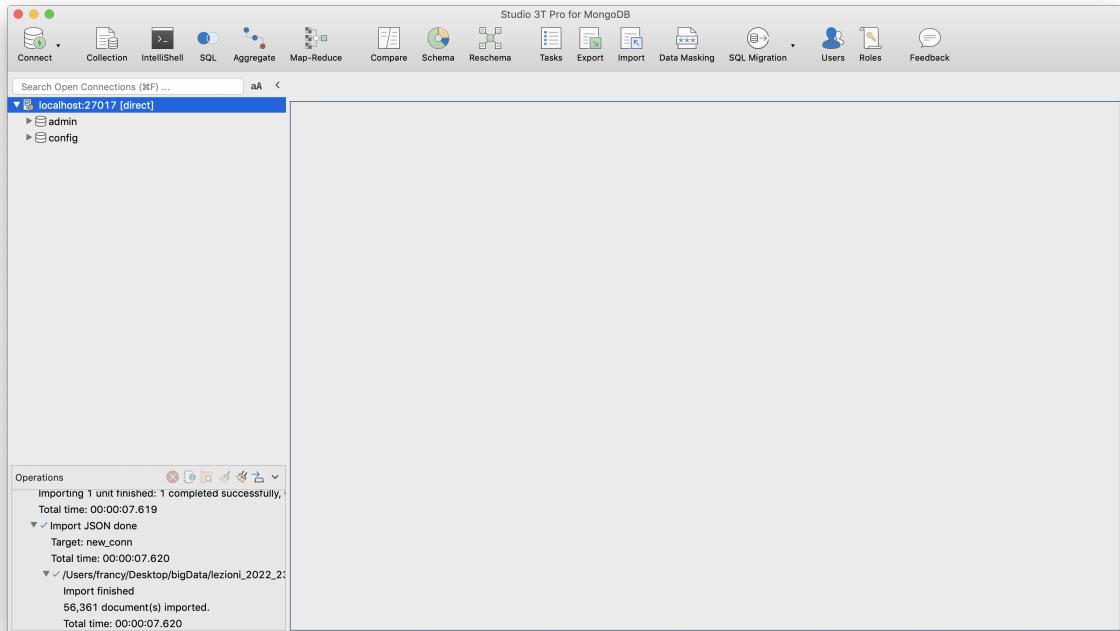
This new form will appear :



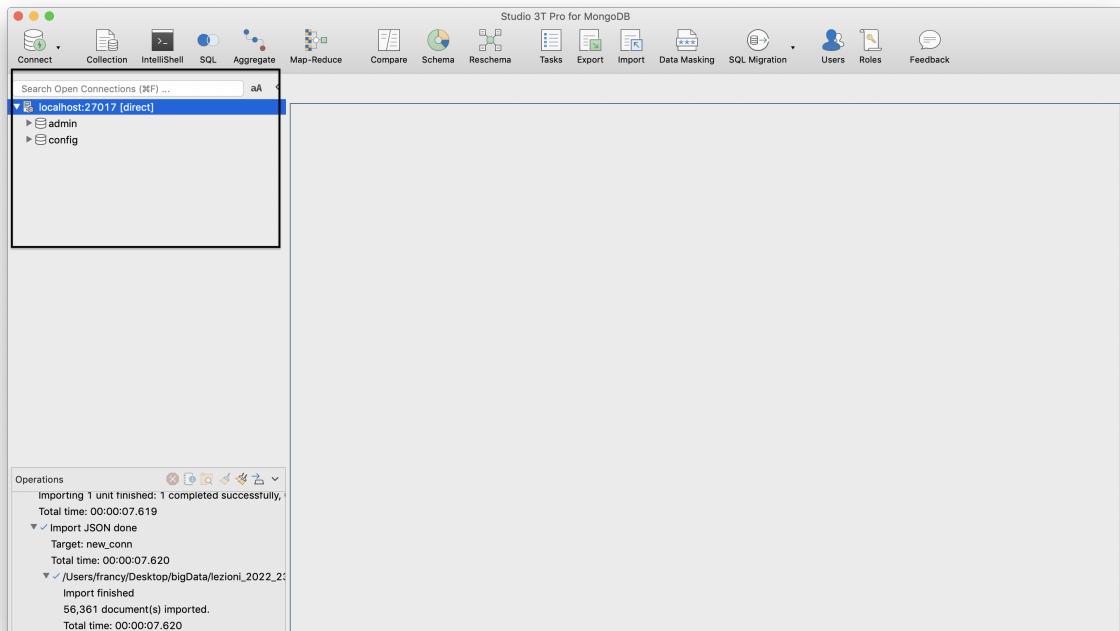
Click on "Connect" :



You will see the Databases appearing on the left side of the screen :



You will see the Databases appearing on the left side of the screen :

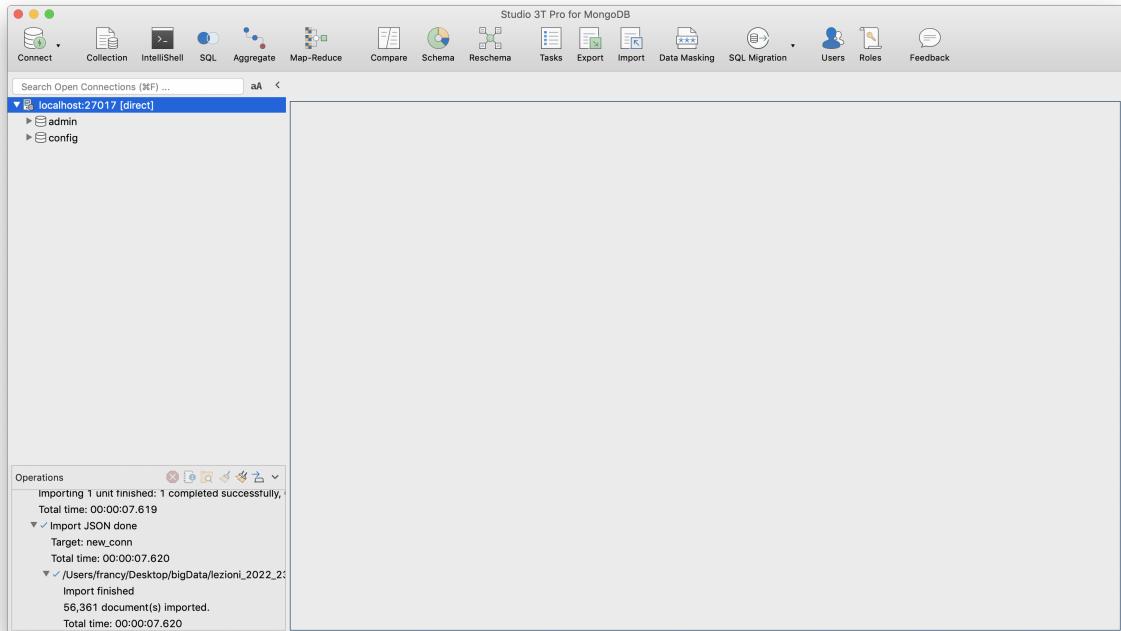


3.2 IMPORT DATA

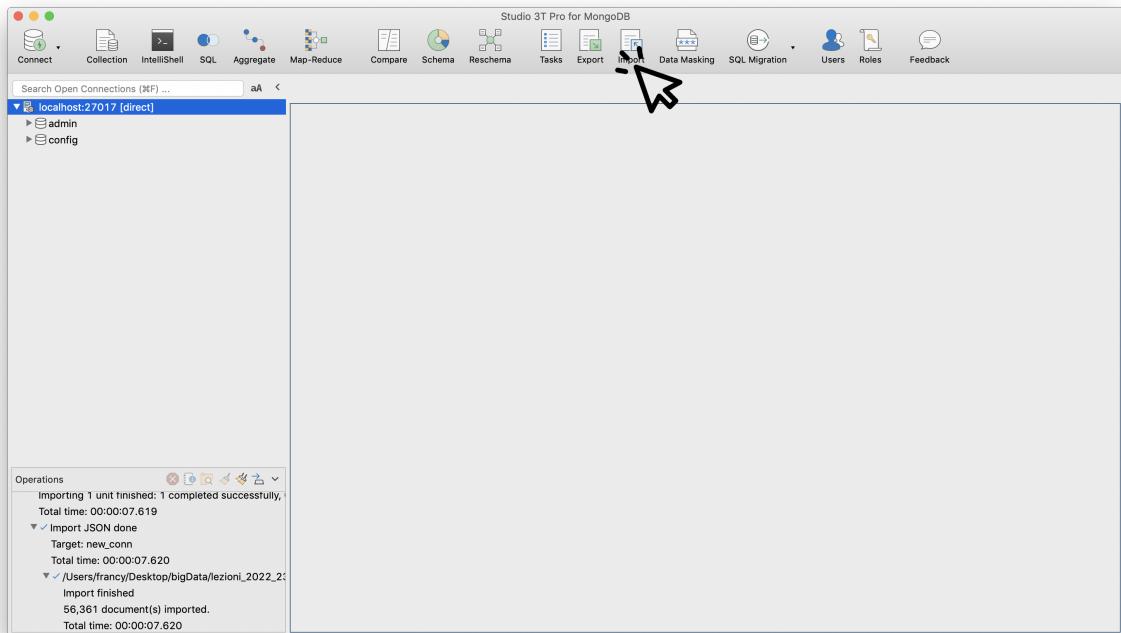
Now you can import data inside the database. Here is the sequence of actions to be run. They will be detailed in the pictures that follow.

- Click on "Import"
- Click on "Configure"
- Click on "Add Source"
- Look for files that you want to import
- Click on "Open"
- Click on "Select Target Connection"
- Select the connection just created and click on "Connect"
- Click on "Run"
- Click on "OK"
- You can see the collections on the left

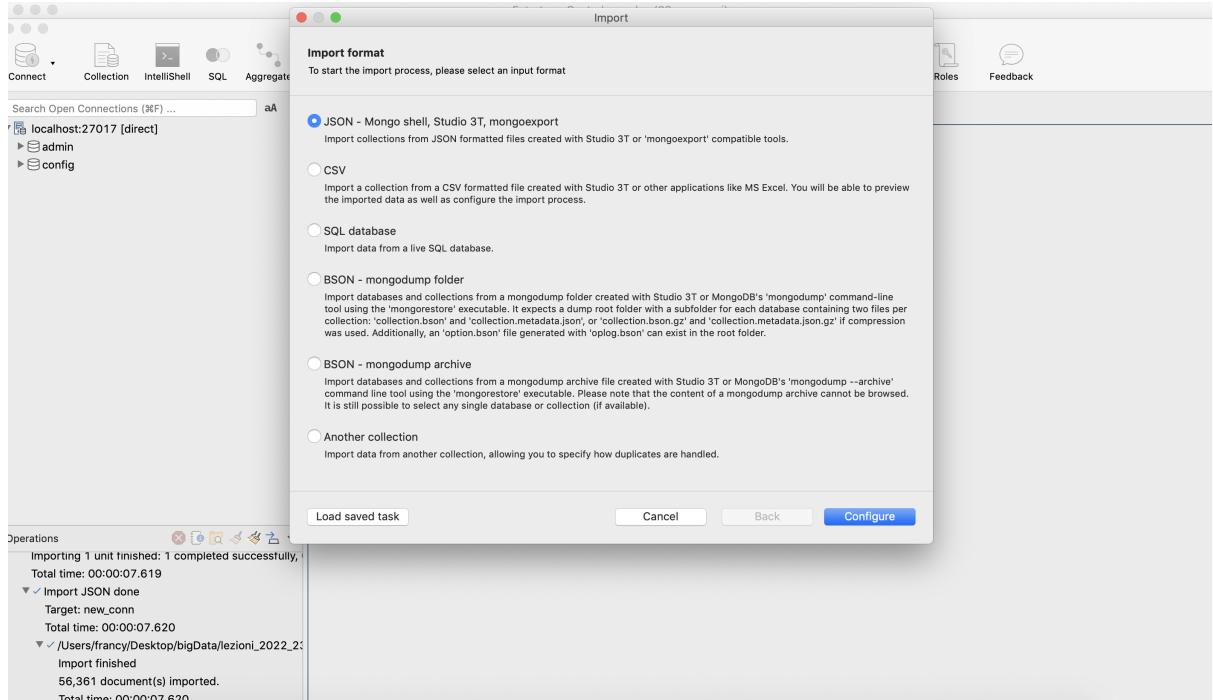
From the main window :



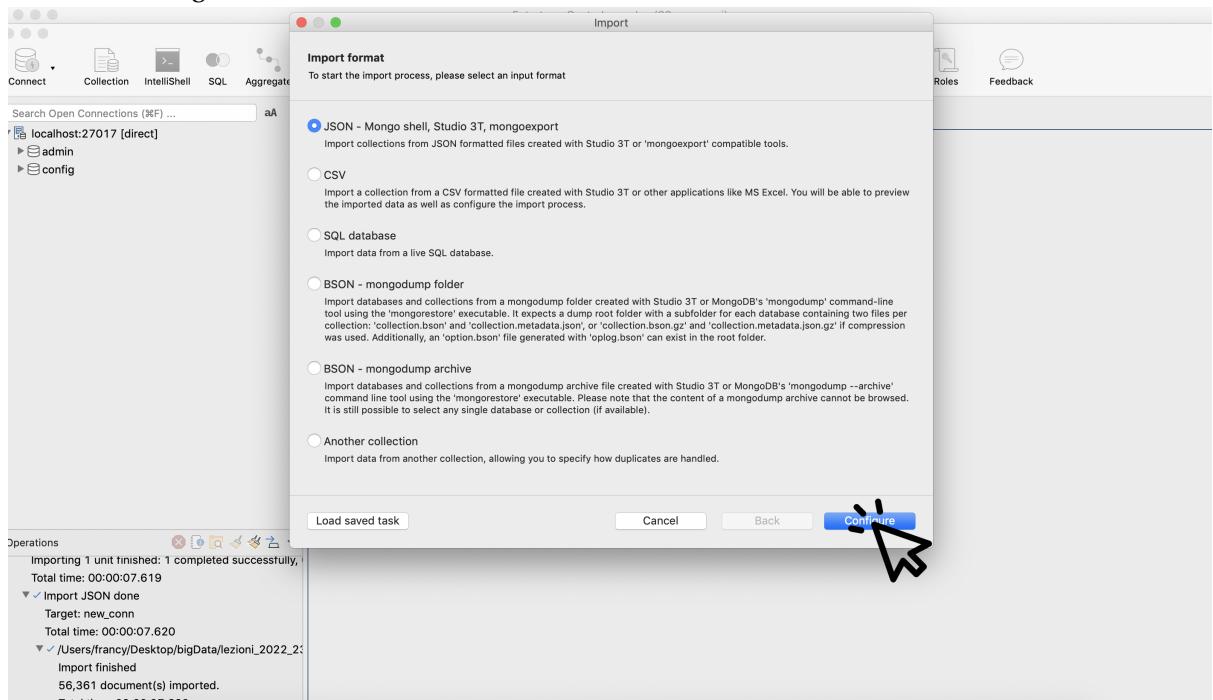
Click on "Import" :



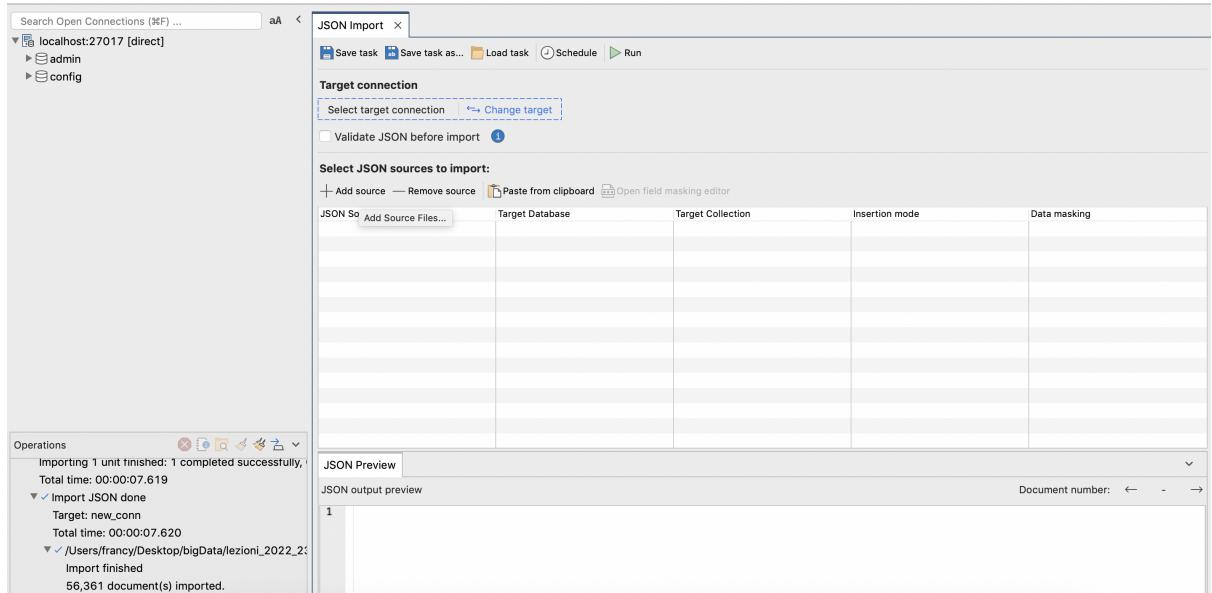
You will see this window :



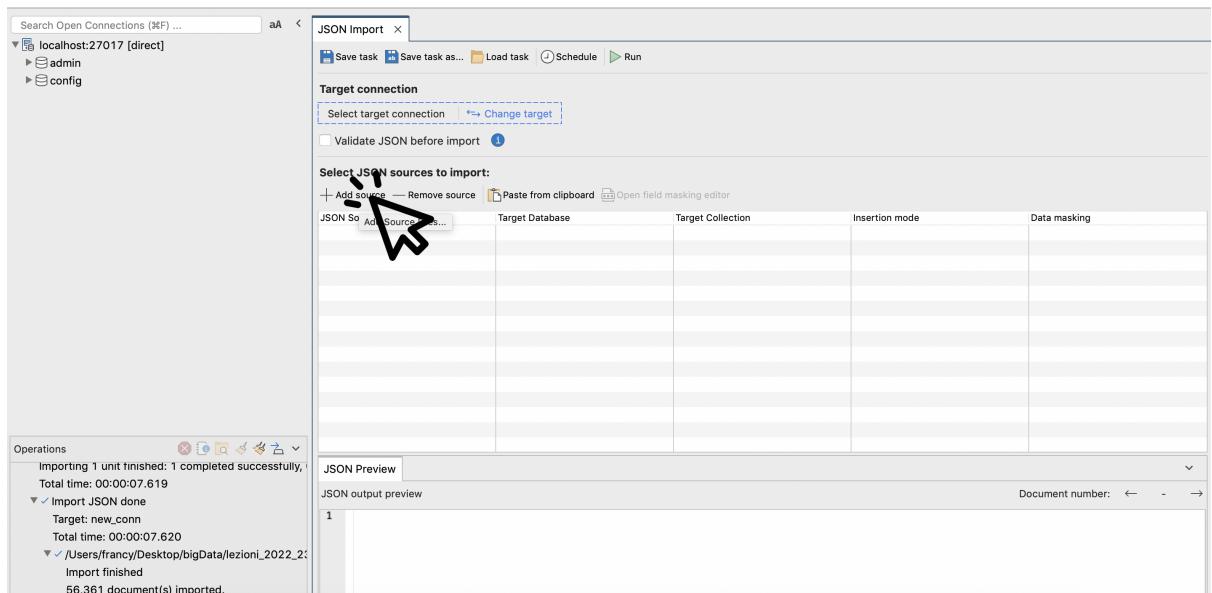
Click on "Configure" :



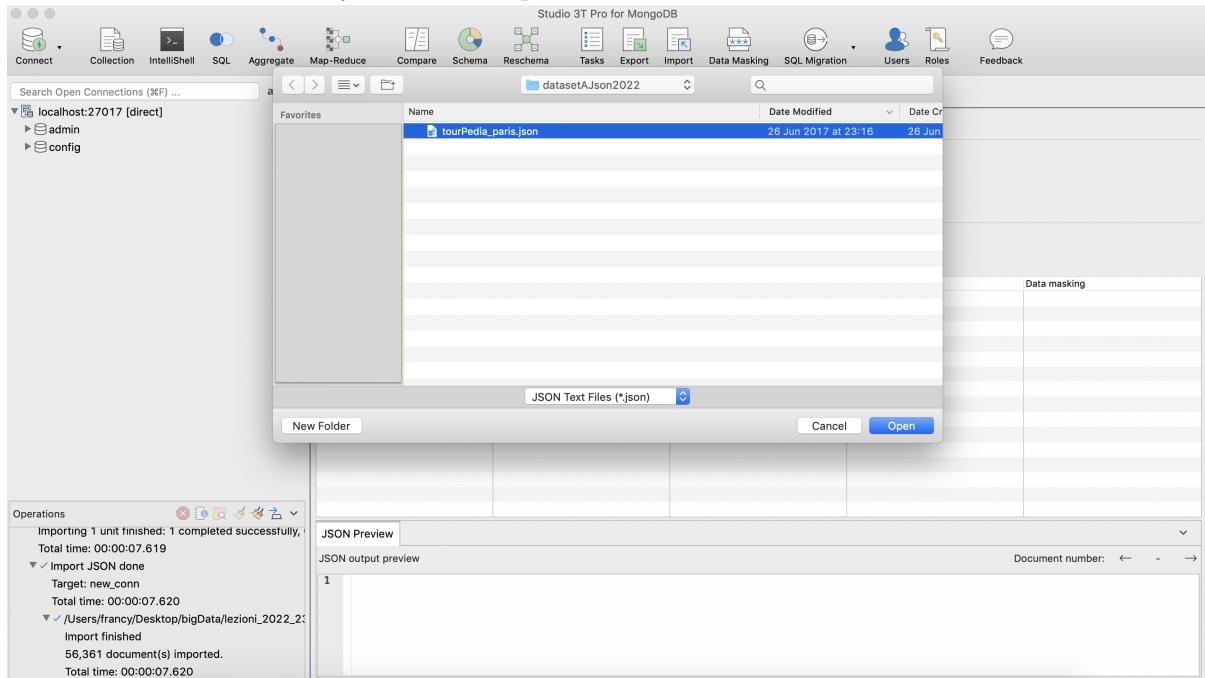
You will see this window :



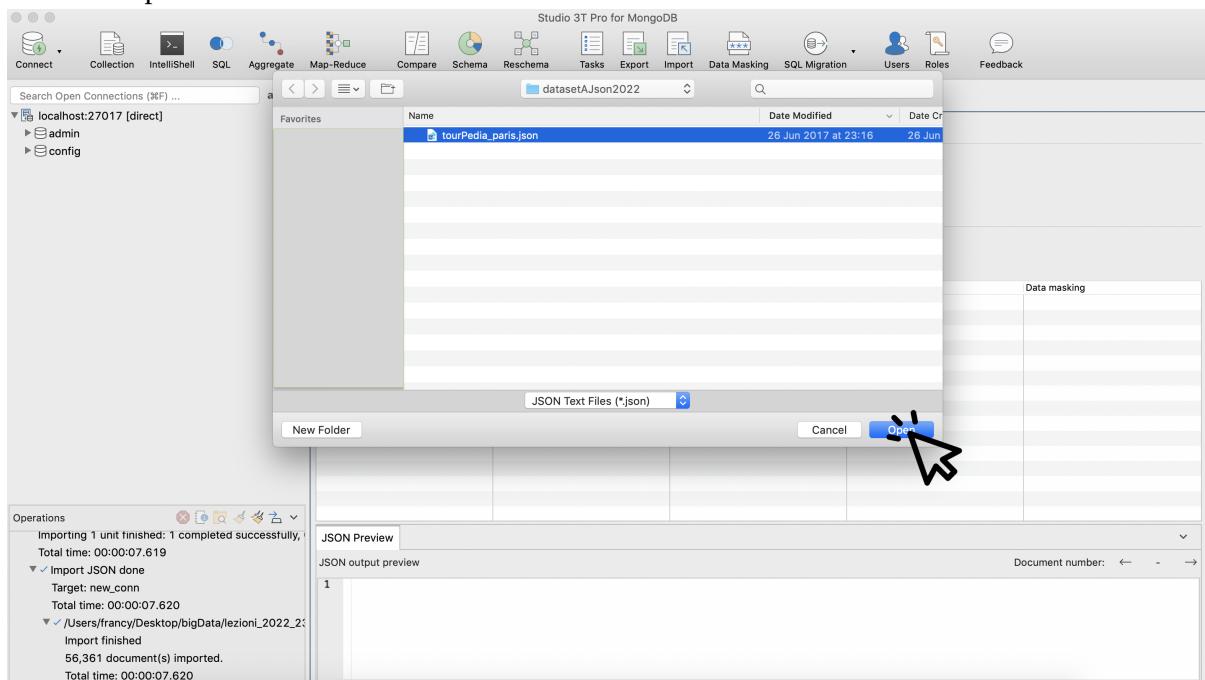
Click on "Add Source" :



Look for the JSON files that you want to import :



Click on "Open" :



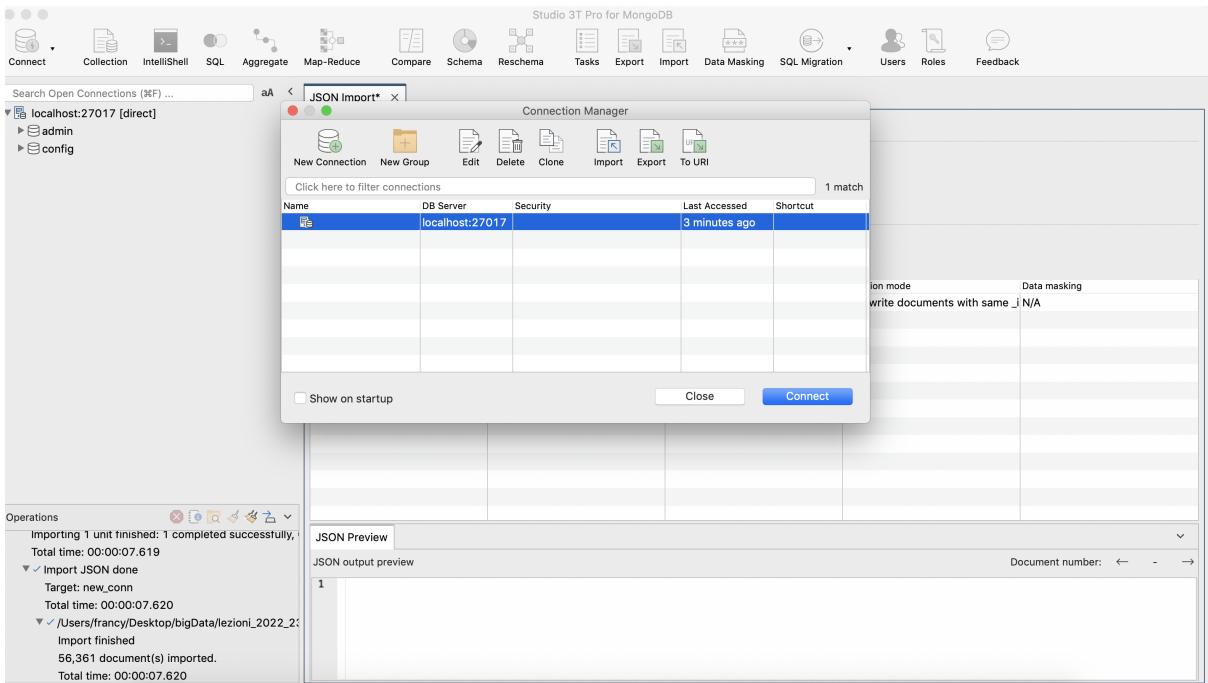
From the same window :

The screenshot shows the MongoDB Compass interface with the "JSON Import" tab selected. In the "Operations" panel on the left, it displays a log entry: "Importing 1 unit finished: 1 completed successfully, Total time: 00:00:07.619". Below this, under "Import JSON done", it shows "Target: new_conn" and "Total time: 00:00:07.620". The main "JSON Import" pane has a "Target connection" section with a "Select target connection" dropdown and a "Change target" button. The "Select JSON sources to import" section lists a single source: "tourPedia_paris.json" with "Target Database" set to "datasetAJson2022" and "Target Collection" set to "tourPedia_paris". The "Insertion mode" is set to "Overwrite documents with same _id N/A". The "JSON Preview" panel at the bottom shows a single document with the number "1".

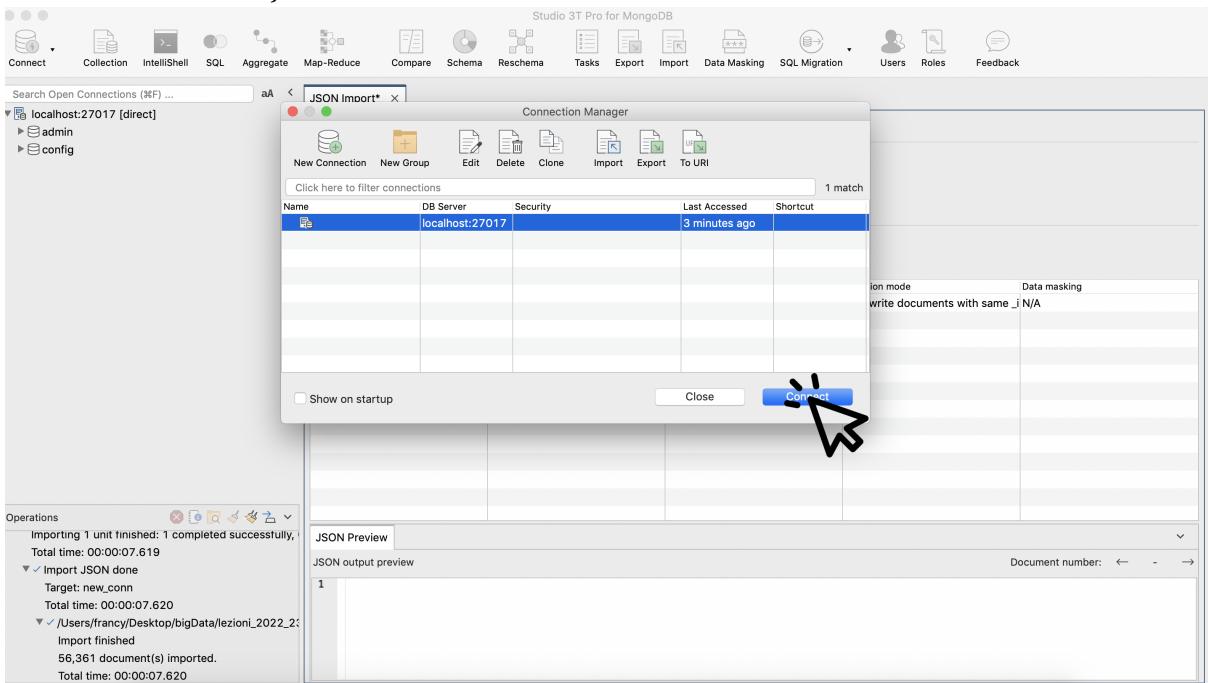
Click on "Select Target Connection" :

This screenshot is identical to the one above, showing the MongoDB Compass "JSON Import" interface. However, a large black arrow points to the "Select target connection" dropdown menu in the "Target connection" section of the main pane. The rest of the interface, including the operations log and preview panel, remains the same.

From the new window :



Select the connection just created and click "Connect" :



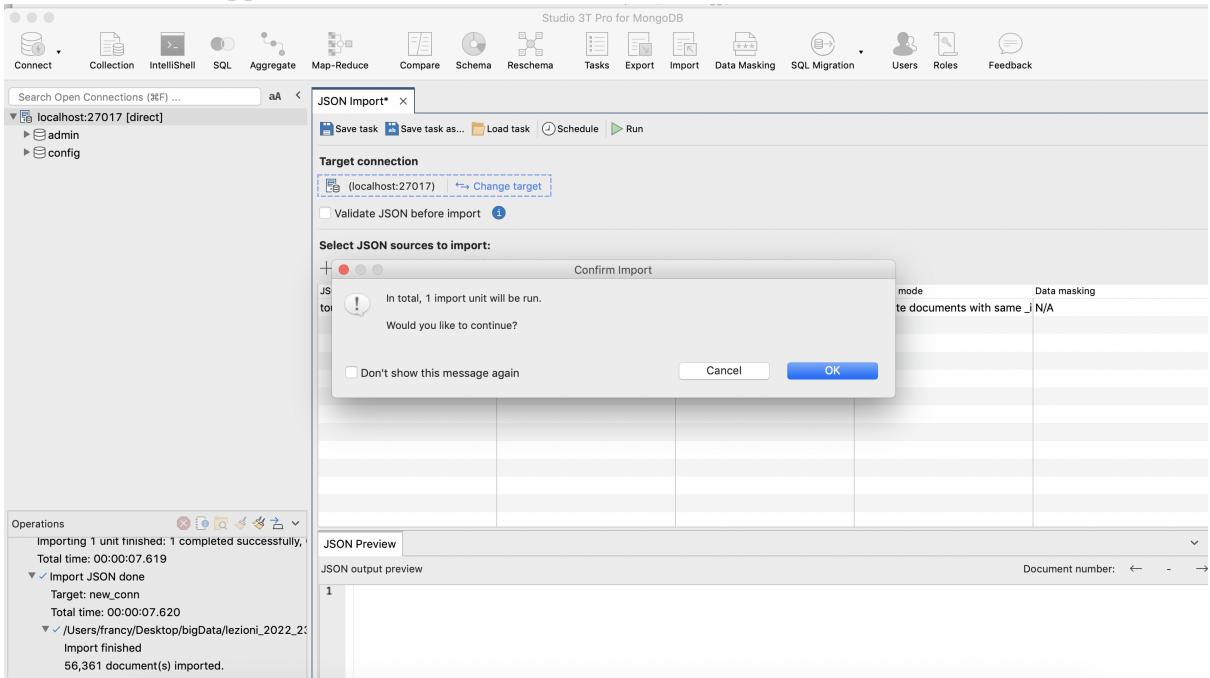
Now all is ready :

The screenshot shows the Studio 3T Pro for MongoDB interface. The main window is titled "JSON Import*" and displays a configuration for importing JSON data. The "Target connection" section shows a target database "(localhost:27017)" and a target collection "tourPedia_paris". The "Select JSON sources to import" section lists a single source "tourPedia_paris.json" with its target database set to "datasetAJson2022" and target collection "tourPedia_paris". The "Operations" panel on the left shows a successful import process: "Importing 1 unit finished: 1 completed successfully, Total time: 00:00:07.619" and "Import JSON done Target: new_conn Total time: 00:00:07.620". The "JSON Preview" panel on the right shows the imported document structure.

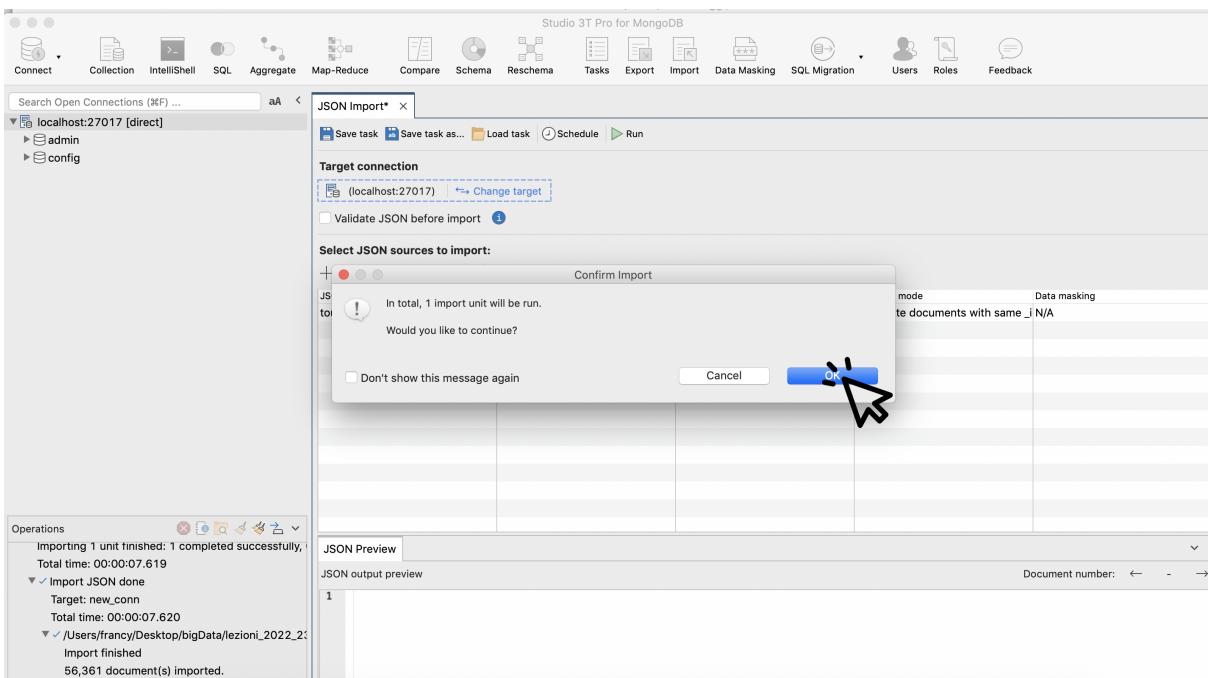
Click on "Run" :

This screenshot is identical to the previous one, showing the "JSON Import*" interface. However, a large black cursor arrow points directly at the "Run" button located in the top right corner of the main window. The "Operations" panel and "JSON Preview" panel are also visible.

This window will appear :



Click on "OK" :



The new database will appear on the left :

JSON Source	Target Database	Target Collection	Insertion mode	Data masking
tourPedia_paris.json	datasetAJson2022	tourPedia_paris	Overwrite documents with same _id N/A	

As you can see the name of the database is the same of the folder in which the json file was stored :

JSON Source	Target Database	Target Collection	Insertion mode	Data masking
tourPedia_paris.json	datasetAJson2022	tourPedia_paris	Overwrite documents with same _id N/A	

If you click on the database :

The screenshot shows the Studio 3T Pro for MongoDB interface. The left sidebar lists databases: admin, config, and datasetAJson2022. The main area is titled "JSON Import*" and shows a table for selecting JSON sources to import. The table has columns: JSON Source, Target Database, Target Collection, Insertion mode, and Data masking. One row is present: tourPedia_paris.json, datasetAJson2022, tourPedia_paris, Overwrite documents with same _id N/A. Below the table is a "JSON Preview" section showing a single document with the number 1. The bottom-left pane shows the "Operations" history with a successful import task.

You can see the collections :

This screenshot is similar to the previous one but with a different selection in the left sidebar: datasetAJson2022 is now expanded, showing its collections: Collections (1), tourPedia_paris; GridFS Buckets (0); System (0); and Views (0). The rest of the interface is identical to the first screenshot, showing the JSON Import dialog and the successful import operation in the operations log.

This screenshot shows the interface after the import process has completed. The left sidebar shows the imported database datasetAJson2022 with its collections listed. The main area still displays the "JSON Import*" dialog, but the "Operations" history pane at the bottom is no longer visible.

3.3 INTELLISHELL

You can open the IntelliShell and run the queries from this interface. Here is the sequence of actions to be run. They will be detailed in the pictures that follow.

- From the main page.
- Click on the connection
- Click on the "IntelliShell"

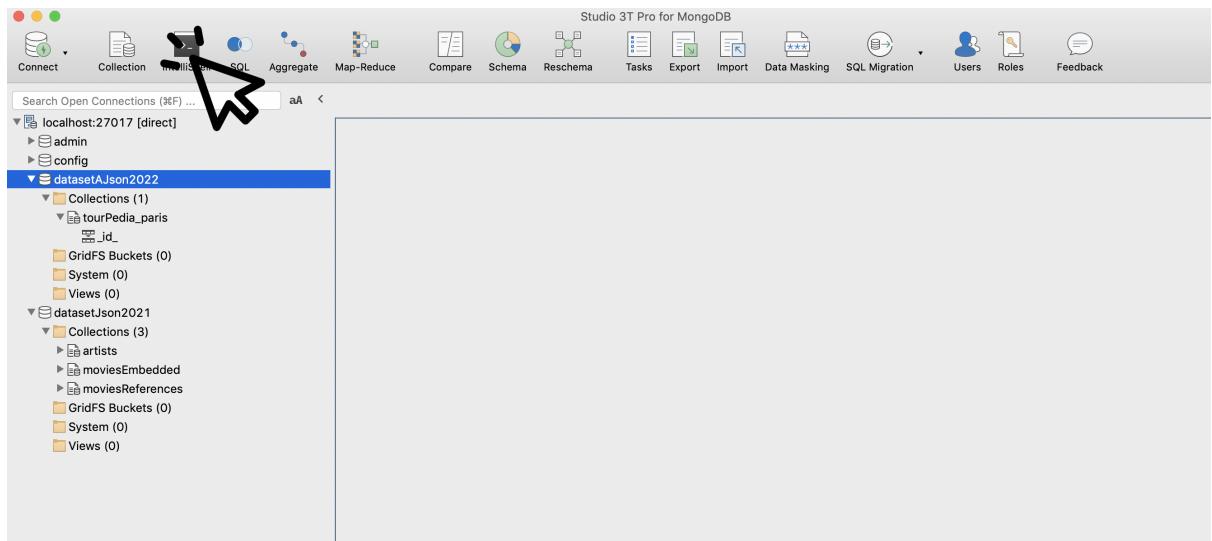
From the main page :

The screenshot shows the Studio 3T Pro for MongoDB interface. At the top, there's a toolbar with various icons: Connect, Collection, IntelliShell, SQL, Aggregate, Map-Reduce, Compare, Schema, Reschema, Tasks, Export, Import, Data Masking, SQL Migration, Users, Roles, and Feedback. Below the toolbar is a search bar labeled "Search Open Connections (%F) ...". The main pane displays a tree view of database connections. Under "localhost:27017 [direct]", there are two databases: "admin" and "config". Under "datasetAJson2022", there is one collection named "tourPedia_paris" and other system-related items. Under "datasetJson2021", there are three collections: "artists", "moviesEmbedded", and "moviesReferences", along with other system-related items. At the bottom left, there's an "Operations" panel showing a successful import operation from November 18, 2022, at 10:02:32. The log message indicates 3 units imported successfully in 0:00:00.363.

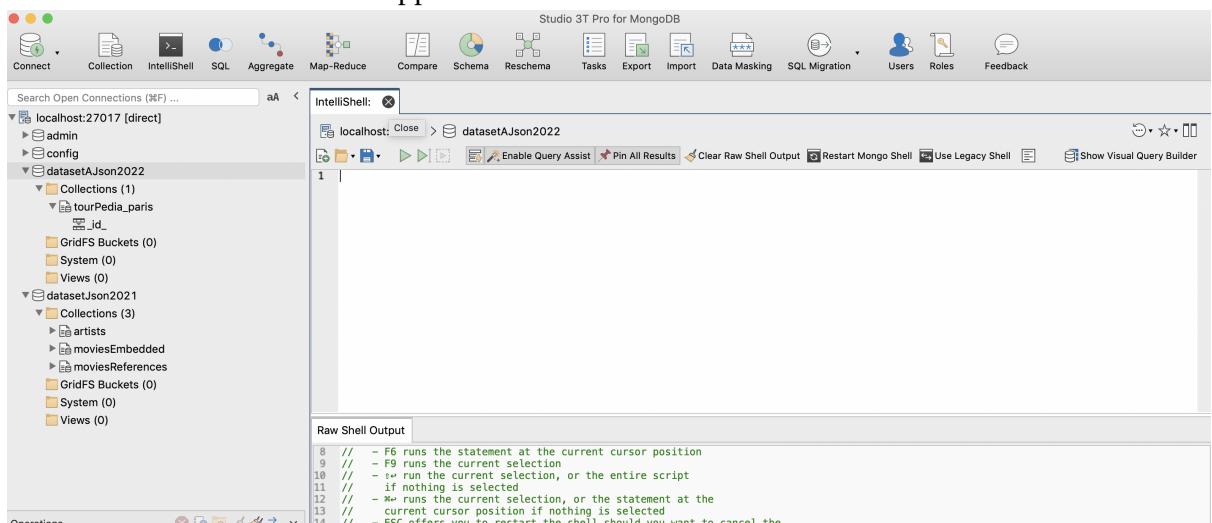
Click on the connection :

This screenshot is similar to the previous one, showing the Studio 3T Pro interface. The main difference is that the "datasetAJson2022" connection is now highlighted with a blue selection bar. The tree view shows the same structure as before, with "datasetAJson2022" expanded to show its collections and system items. The "Operations" panel is not visible in this specific screenshot.

Click on the IntelliShell :



You can see the IntelliShell will appear :



You can write there your queries :

```

IntelliShell: x
localhost:27017 > admin
localhost:27017 > datasetAJson2022
1 o show collection
      show collections

Raw Shell Output
8 // - F6 runs the statement at the current cursor position
9 // - F9 runs the current selection
10 // - Esc runs the current selection, or the entire script
11 // - if nothing is selected
12 // - M+ runs the current selection, or the statement at the
13 //   current cursor position if nothing is selected
14 // - ESC offers you to restart the shell should you want to cancel the
15 //   execution of the current command or script
16 // - "cls" clears the Raw Shell Output tab
17 // - Run Shell Queries via Studio 3T for full access to query results
18 //

Current Mongosh Log ID: 637765d638f4e05936971484
Using Mongosh: 1.5.4
For mongosh info see: https://docs.mongodb.com/mongodb-shell/
20
21
22
23
24
25 >
26 > admin
27
28

```

And see the results on the bottom :

```

IntelliShell: x
localhost:27017 > datasetAJson2022
localhost:27017 > datasetAJson2022
1 show collections

Raw Shell Output
10 // - !> run the current selection, or the entire script
11 // - if nothing is selected
12 // - M+ runs the current selection, or the statement at the
13 //   current cursor position if nothing is selected
14 // - ESC offers you to restart the shell should you want to cancel the
15 //   execution of the current command or script
16 // - "cls" clears the Raw Shell Output tab
17 // - Run Shell Queries via Studio 3T for full access to query results
18 //

Current Mongosh Log ID: 63776b765ee00a85620ac398
Using Mongosh: 1.5.4
For mongosh info see: https://docs.mongodb.com/mongodb-shell/
20
21
22
23
24
25 > datasetAJson2022
26 tourPedia_paris
27 tourPedia_paris
28 tourPedia_paris

```

You can also navigate the collections :

And the nested fields :

4 QUERYING DATA

Now that you are familiar with the environment and the MongoDB query syntax find the queries to retrieve the following data :

1. The movies titled `Gladiator`.
2. The distinct genre values of movies.
3. The movies of `crime` or `drama` genre.
4. The movies released between 1967 and 1995.
5. The list of the movies released between 1967 and 1995, by displaying only title, year, director's last name sorted by year.
6. The same query by adding actor's last name.
7. The number of movies by country.
8. The number of movies by country and by year.
9. The number of movie of genre `drama` by country.
10. The number of movies by country and director's last name.
11. The number of movies by actor and the actor's last name.
12. The movies in which the artist `John Travolta` has played a role.
13. The number of movies in which the artist `John Travolta` has played a role.

5 USING MANUAL REFERENCES

Import a second dataset that codify the same data using references as you did for the previous datasets. You must import two files :

- `artists.json`
- `moviesReferences.json`

Refresh now the view. You will see that you have two new available collections.

Now we try to interact with the data. Explore them running the `find` operation.

1. In the collection `moviesReferences` how are the directors stored?
2. In the collection `moviesReferences` how are the actors stored?

3. Write the query that returns all the movies directed by Hitchcock as new documents containing all the data of the director in a separate field.

6 MONGODB IN THE CLOUD

Now that you are familiar with MongoDB and how NoSQL databases works you can try to use MongoDB on a cloud environment and put together what we have seen about data distribution and what we have seen about Databases.

6.1 MAIN OBJECTIVE

In this part of the exercise class you can try to set-up your own cluster and see how to use MongoDB instance in the cluster from your machine.

6.2 ATLAS

MongoDB Atlas is a multi-cloud database service provided by the same people that build MongoDB. MongoDB Atlas makes it easy to deploy and manage databases on-demand with MongoDB on AWS, Azure, and Google Cloud. You can access the service and having a more deep look to what is provided using the following link :

Atlas

6.3 IMPORTANT

In the set-up of the machines and of the cluster you ALWAYS have to select the free options. For running this class IS NOT required to buy any subscription or set-up any payment in the connection to the cluster.

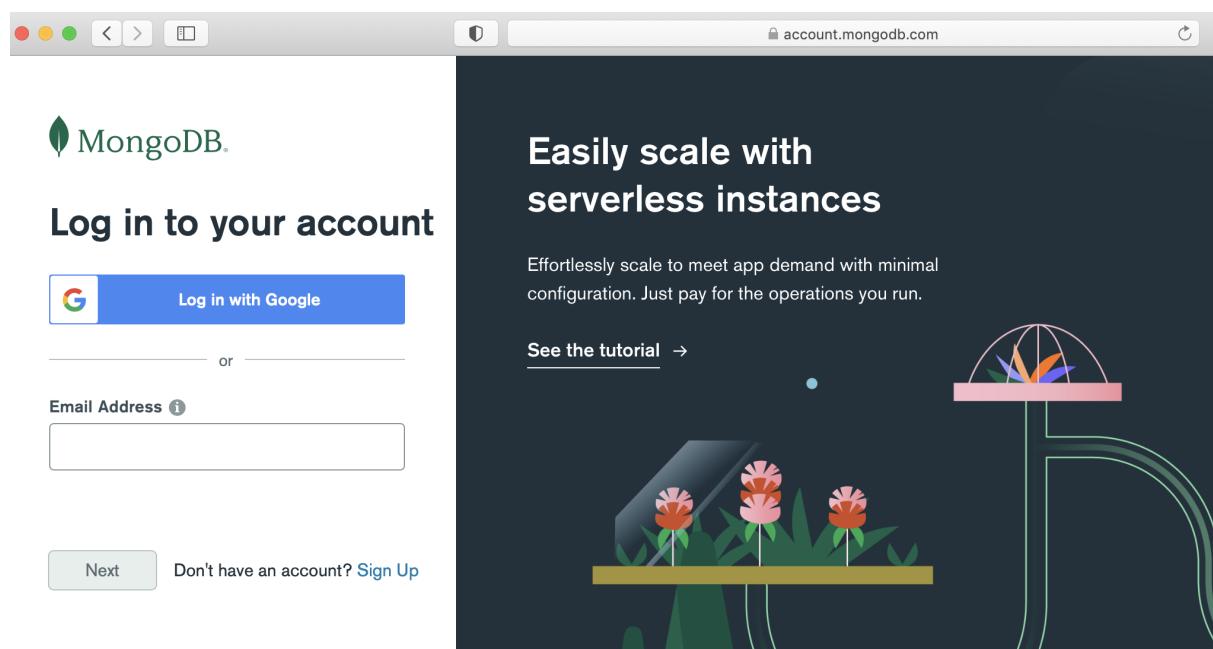
- you will lunch a free cluster
- you will choose your provider and region
- you will run your database
- you will query from your machine.

6.4 CREATE AN ATLAS ACCOUNT OR LOGIN

As first step you must log-in to Atlas and for this step you can use the same credentials you used for downloading your MongoDB community Version.

<https://cloud.mongodb.com>

The fist step will be to login to the system as shown in the following picture.



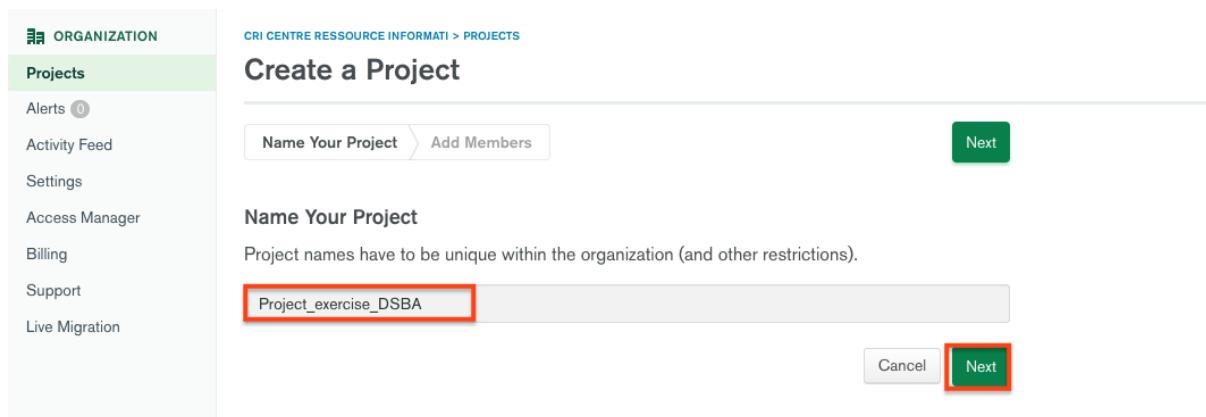
6.5 DEPLOY A FREE CLUSTER

Atlas free clusters provide a development environment to host your data and provide access to a subset of Atlas features that are sufficient for this class.

In order to deploy your cluster you must apply the following steps.

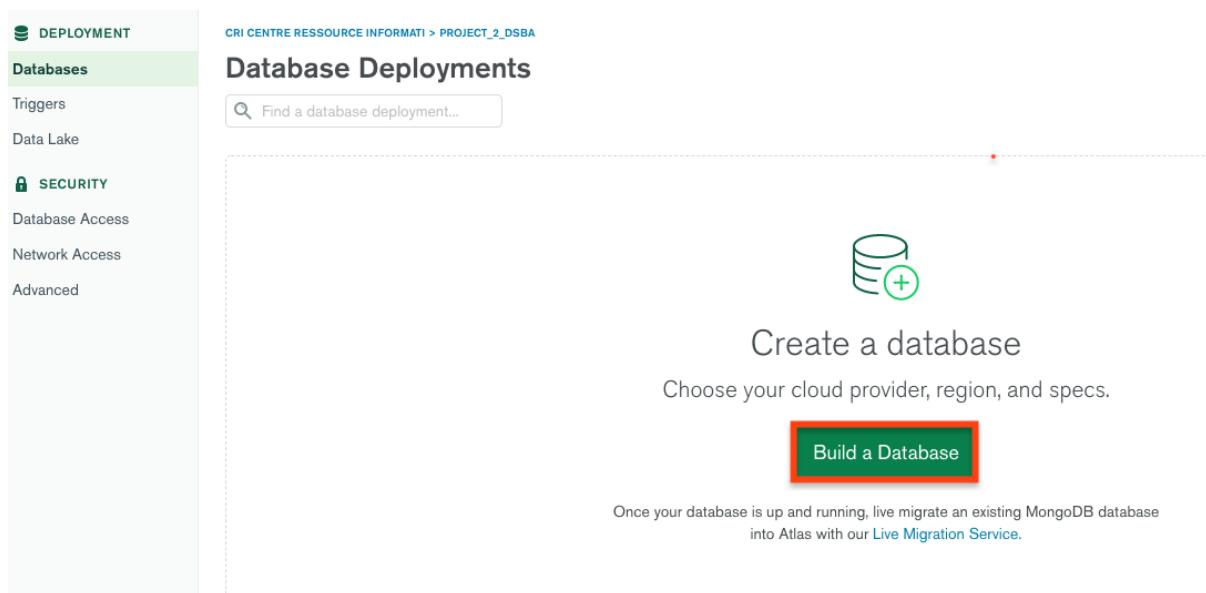
1) Create a new project.

Include the **Name**, in this example **Project_Exercise_DSBA** :



The screenshot shows the 'Create a Project' step of the CRI Centre Ressource Informatique interface. On the left, a sidebar menu includes 'ORGANIZATION' (selected), 'Projects' (highlighted in green), 'Alerts (0)', 'Activity Feed', 'Settings', 'Access Manager', 'Billing', 'Support', and 'Live Migration'. The main area is titled 'Create a Project' with sub-titles 'Name Your Project' and 'Add Members'. A text input field contains 'Project_exercise_DSBA', which is highlighted with a red box. To the right of the input field are 'Next' and 'Cancel' buttons, with 'Next' also highlighted with a red box.

2) Select to deploy a cluster for MongoDB in the following window :



The screenshot shows the 'Database Deployments' step of the CRI Centre Ressource Informatique interface. On the left, a sidebar menu includes 'DEPLOYMENT' (selected), 'Databases' (highlighted in green), 'Triggers', 'Data Lake', 'SECURITY' (with sub-options 'Database Access', 'Network Access', and 'Advanced'), and 'Advanced'. The main area is titled 'Database Deployments' with a search bar 'Find a database deployment...'. It features a large 'Create a database' button with a cylinder icon containing a plus sign, which is highlighted with a red box. Below the button, text reads 'Choose your cloud provider, region, and specs.' and 'Once your database is up and running, live migrate an existing MongoDB database into Atlas with our [Live Migration Service](#)'. A 'Build a Database' button is located at the bottom of the main area.

3) Select **Shared** Clusters and click **Create a Cluster** :

The screenshot shows the MongoDB deployment options page. At the top, a message reads: "Experience the best of MongoDB on AWS, Azure, and Google Cloud. Choose a deployment option to get started." Below this are three options:

- Serverless**: A white card with a green icon. Description: "For serverless applications that aren't critical with variable traffic. Minimal configuration required." Benefits: Pay only for operations run, resources scale seamlessly, always-on security. Price: Starting at \$0.30/1M reads.
- Dedicated**: A white card with a green icon. Description: "For production applications with sophisticated workload requirements. Advanced configuration controls." Benefits: Network isolation, fine-grained access controls, on-demand performance advice, multi-region/multi-cloud options. Price: Starting at \$0.08/hr*, estimated cost \$56.94/month.
- Shared**: A white card with a green icon, highlighted with a red border. Description: "For learning and exploring MongoDB in a cloud environment. Basic configuration options." Benefits: No credit card required, explore with sample datasets, upgrade to dedicated clusters. Price: FREE.

In the bottom right corner of the Shared card, there is a large green "Create" button.

4) Select your preferred Cloud Provider & Region and clic Create Cluster :

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

PREVIEW Serverless Dedicated FREE Shared

Do not change the default options for the Cluster Tier (MO Sandbox) and do not select additional settings.

5) Optionally change the cluster name :

Cluster Tier	M0 Sandbox (Shared RAM, 512 MB Storage) Encrypted
Additional Settings	MongoDB 4.4, No Backup
Cluster Name	<div style="border: 2px solid red; padding: 10px;"><div style="border: 2px solid blue; padding: 5px; margin-bottom: 10px;">ClusterLabDSBA</div><p>Cluster names can only contain ASCII letters, numbers, and hyphens.</p></div>

6) Finally click on Create Cluster :



The deployment will start :

Database Deployments

VERSION	REGION	CLUSTER TIER	TYPE	BACKUPS	LINKED REALM APP	ATLAS SEARCH
4.4.10	AWS / N. Virginia (us-east-1)	M0 Sandbox (General)	Replica Set - 3 nodes	Inactive	None Linked	Create Index

And finally you will have your cluster running as shown in this picture :

Database Deployments

VERSION	REGION	CLUSTER TIER	TYPE	BACKUPS	LINKED REALM APP	ATLAS SEARCH
4.4.10	AWS / N. Virginia (us-east-1)	M0 Sandbox (General)	Replica Set - 3 nodes	Inactive	None Linked	Create Index

You can start looking at this interface and see the concepts we have already seen at lesson : the region, the replica set (equal 3 by default).

6.6 TO CONFIGURE AN IP ADDRESS

At this point you must configure how and who can access the cluster from the outside. In Atlas, you can only connect to a cluster only from a trusted IP address. Then you must create a list of trusted IP addresses, referred to as a IP access list, that can be used to connect to your cluster and access your data.

1) Click **Connect** in the top-left corner of Atlas :

The screenshot shows the MongoDB Atlas interface. On the left, there's a sidebar with sections for Deployment, Databases, Triggers, Data Lake, Security, Database Access, and Advanced. The main area is titled "Database Deployments" and shows a single deployment named "ClusterLabDSBA". Below the deployment name is a "Connect" button, which is highlighted with a red box. Other buttons include "View Monitoring", "Browse Collections", and "...". To the right of the deployment details, there are tabs for "FREE" and "SHARED". Below the deployment list, there's a section for "Enhance Your Experience" with a "Upgrade" button. At the bottom, there are filters for VERSION (4.4.10), REGION (AWS / N. Virginia (us-east-1)), CLUSTER TIER (M0 Sandbox (General)), TYPE (Replica Set - 3 nodes), BACKUPS (Inactive), LINKED REALM APP (None Linked), and ATLAS SEARCH (Create Index).

2) Click **Add Your Current IP Address** :

The screenshot shows the "Add a connection IP address" step. It has three buttons: "Add Your Current IP Address" (highlighted with a red box), "Add a Different IP Address", and "Allow Access from Anywhere". Below this, there's a section for creating a database user. It says "This first user will have atlasAdmin permissions for this project." and "Keep your credentials handy, you'll need them for the next step." It includes fields for "Username" (ex. dbUser) and "Password" (ex. dbUserPassword). There's also an "Autogenerate Secure Password" checkbox and a "SHOW" button. A "Create Database User" button is at the bottom. At the very bottom, there are "Close" and "Choose a connection method" buttons.

Your IP address will be automatically discovered and you can add it clicking on **Add IP address**.

① Add a connection IP address

The dialog box has two main sections: 'IP Address' and 'Description (Optional)'. The 'IP Address' section contains a text input field with the value '237.234.456.22'. The 'Description (Optional)' section contains a text input field with the placeholder 'An optional comment describing this entry'. At the bottom right are two buttons: 'Cancel' and a green 'Add IP Address' button, which is highlighted with a red border.

After you confirm clicking on **Add IP Address** you will have the following feedback :

You need to secure your MongoDB Atlas cluster before you can use it. Set which users and IP addresses can access your cluster now. [Read more ↗](#)

You can't connect yet. Set up your user security permission below.

① Add a connection IP address

✓ An IP address has been added to the IP Access List. [Add another address in the IP Access List tab.](#)

② Create a Database User

This first user will have **atlasAdmin ↗** permissions for this project.

Keep your credentials handy, you'll need them for the next step.

Username

ex. dbUser

Password

ex. dbUserPassword

Autogenerate Secure Password

SHOW

Create Database User

Close

Choose a connection method

6.7 ADD A DATABASE USER

You must create a database user to access your cluster. For security purposes, Atlas requires clients to authenticate as MongoDB database users to access clusters. In this step you must create a database user with username and password.

Add your data into the form and confirm by clicking on **Create database user**:

2 Create a Database User

This first user will have [atlasAdmin](#) permissions for this project.
Keep your credentials handy, you'll need them for the next step.

Username	Password	Autogenerate Secure Password
franbugiotti	*****	SHOW

Create Database User

After you confirm clicking on **Create database user** you will have the following feedback:

Connect to ClusterLabDSBA

Setup connection security > Choose a connection method > Connect

You need to secure your MongoDB Atlas cluster before you can use it. Set which users and IP addresses can access your cluster now. [Read more](#)

You're ready to connect. Choose how you want to connect in the next step.

1 Add a connection IP address

- ✓ An IP address has been added to the IP Access List. [Add another address in the IP Access List tab.](#)

2 Create a Database User

- ✓ A MongoDB user has been added to this project. [Not yours? Create one in the MongoDB Users tab.](#)

You'll need your MongoDB user's credentials in the next step.

Close

Choose a connection method

6.8 DATABASE CONNECTION

You can create a database user to access your Atlas database deployment in the Connect dialog or you can also add the database user from the Database Deployment view following the previous steps and clicking on the option Choose a connection method.

In the code of your application shows :

- The <username> of the user that will connect to the database
- The string <password> to be replaced **in the code** with the password specified when you created your database user.
- Replace myFirstDatabase with the name of the database that connections will use by default.
Notice that if you omit the database, the test database is used by default.

- 1) Select your driver and version from the dropdown menus. The code sample containing a connection string will display.

In the following we selected Python as language and 3.6 or later as version :

Connect to ClusterLabDSBA

✓ Setup connection security ✓ Choose a connection method Connect

1 Select your driver and version

DRIVER	VERSION
Python	3.6 or later

2 Add your connection string into your application code

Include full driver code example

```
mongodb+srv://franbugiotti:  
<password>@clusterlabdsba.jgafb.mongodb.net/myFirstDatabase?  
retryWrites=true&w=majority
```



Replace <password> with the password for the **franbugiotti** user. Replace **myFirstDatabase** with the name of the database that connections will use by default. Ensure any option params are **URL encoded**.

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

The <username> to be changed in your code :

Connect to ClusterLabDSBA

✓ Setup connection security ✓ Choose a connection method Connect

1 Select your driver and version

DRIVER	VERSION
Python	3.6 or later

2 Add your connection string into your application code

Include full driver code example

```
mongodb+srv://<username>:  
<password>@clusterlabdsba.jgafb.mongodb.net/myFirstDatabase?  
retryWrites=true&w=majority
```

Replace <password> with the password for the <username> user. Replace myFirstDatabase with the name of the database that connections will use by default. Ensure any option params are [URL encoded](#).

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

The <password> string to be replaced in your code :

Connect to ClusterLabDSBA

✓ Setup connection security ✓ Choose a connection method Connect

1 Select your driver and version

DRIVER	VERSION
Python	3.6 or later

2 Add your connection string into your application code

Include full driver code example

```
mongodb+srv://<username>:  
<password>@clusterlabdsba.jgafb.mongodb.net/myFirstDatabase?  
retryWrites=true&w=majority
```

Replace <password> with the password for the <username> user. Replace myFirstDatabase with the name of the database that connections will use by default. Ensure any option params are [URL encoded](#).

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

Go Back

Close

The database to be connected to (to be replaced with your target) :

Connect to ClusterLabDSBA

✓ Setup connection security ✓ Choose a connection method Connect

① Select your driver and version

DRIVER	VERSION
Python	3.6 or later

② Add your connection string into your application code

Include full driver code example

```
mongodb+srv://<username>:  
<password>@clusterlabdsba.jgafb.mongodb.net/myFirstDatabase?  
retryWrites=true&w=majority
```



Replace **<password>** with the password for the **<username>** user. Replace **myFirstDatabase** with the name of the database that connections will use by default. Ensure any option params are **URL encoded**.

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

[Go Back](#)

[Close](#)

6.9 LOAD SAMPLE DATA

Atlas provides sample data you can load into your Atlas clusters. Otherwise you can use the **Migrate or Import Data** option.

We suggest you to start with sample data and you can perform this import clicking on **Load Sample Dataset**:

The screenshot shows the MongoDB Atlas interface with the 'ClusterLabDSBA' cluster selected. The 'Collections' tab is active. At the bottom left, there is a green button labeled 'Load a Sample Dataset' which is highlighted with a red box. To its right is another button labeled 'Add My Own Data'. Below these buttons is a link 'Learn more in Docs and Tutorials'.

Here the default database list after the import :

The screenshot shows the MongoDB Atlas interface with the 'ClusterLabDSBA' cluster selected. The 'Collections' tab is active. On the left, there is a sidebar titled 'NAMESPACES' with a list of databases and collections. One database, 'sample_airbnb', is highlighted with a red box. The main area displays the 'sample_airbnb.listingsAndReviews' collection. It shows the collection size (89.99MB), total documents (6555), and indexes total size (516KB). Below this, there are tabs for 'Find', 'Indexes', 'Schema Anti-Patterns', 'Aggregation', and 'Search Indexes'. A 'FILTER' button is present. The 'QUERY RESULTS 1-20 OF MANY' section shows a single document with various fields like '_id', 'listing_url', 'name', 'summary', etc. At the bottom, there is a link 'SHOW 14 MORE FIELDS'.

And the **sample_supplies** to which we connect with the provided notebook in the following :

CRI CENTRE RESSOURCE INFORMATI > PROJECT_LAB_DSBA > DATABASES

ClusterLabDSBA

Overview Real Time Metrics **Collections** Search Profiler

DATABASES: 8 COLLECTIONS: 21

+ Create Database

NAMESPACES

- ▶ sample_airbnb
- ▶ sample_analytics
- ▶ sample_geospatial
- ▶ sample_mflix
- ▶ sample_restaurants
- sample_supplies** +
- | sales
- ▶ sample_training
- ▶ sample_weatherdata

sample_supplies.sales

COLLECTION SIZE: 4.13MB TOTAL DOCUMENTS: 5000 INC

Find Indexes Schema Anti-Patterns 0

FILTER { field: 'value' }

QUERY RESULTS 1-20 OF MANY

```
_id: ObjectId("5bd761dcae323e45a93ccfe8")
saleDate: 2015-03-23T21:06:49.506+00:00
> items: Array
  storeLocation: "Denver"
> customer: Object
  couponUsed: true
  purchaseMethod: "Online"

_id: ObjectId("5bd761dcae323e45a93ccfe9")
saleDate: 2015-08-25T10:01:02.918+00:00
> items: Array
  storeLocation: "Seattle"
> customer: Object
  couponUsed: false
  purchaseMethod: "Phone"
```

Notice that you can run simple queries using filters in this form :

Find Indexes Schema Anti-Patterns 0 Aggregation Search Indexes 0

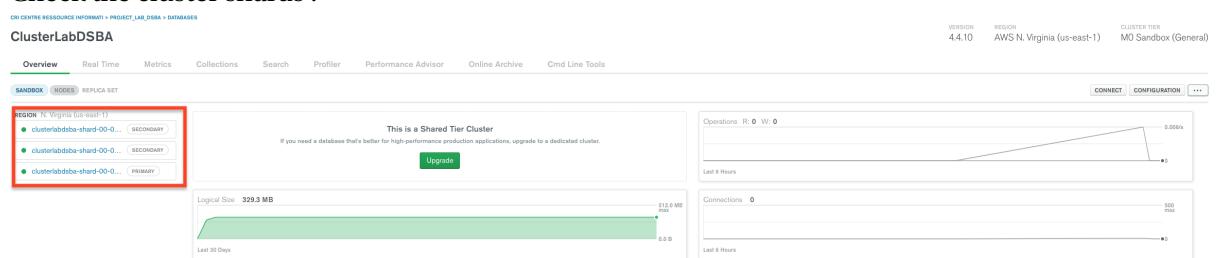
FILTER { field: 'value' }

QUERY RESULTS 1-20 OF MANY

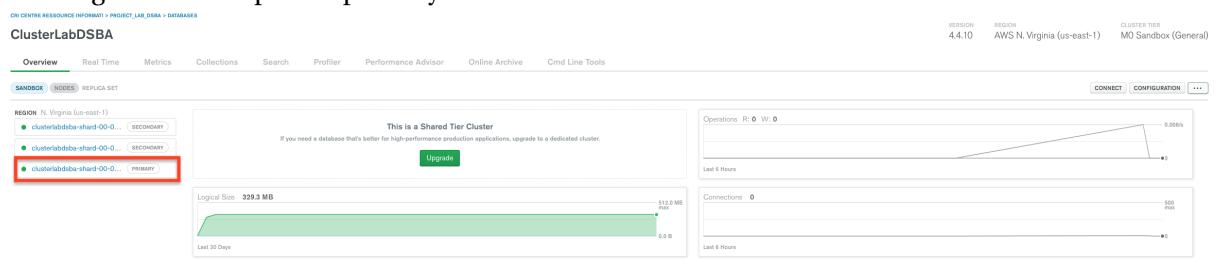
6.10 LOOK AT THE DASHBOARD

You can have a look at the dash board and set-up some considerations about the cluster usage. A strong suggestion is to check the dashboard right now and verify how the cluster react to write and read queries. A write query is performed when you import your dataset for example.

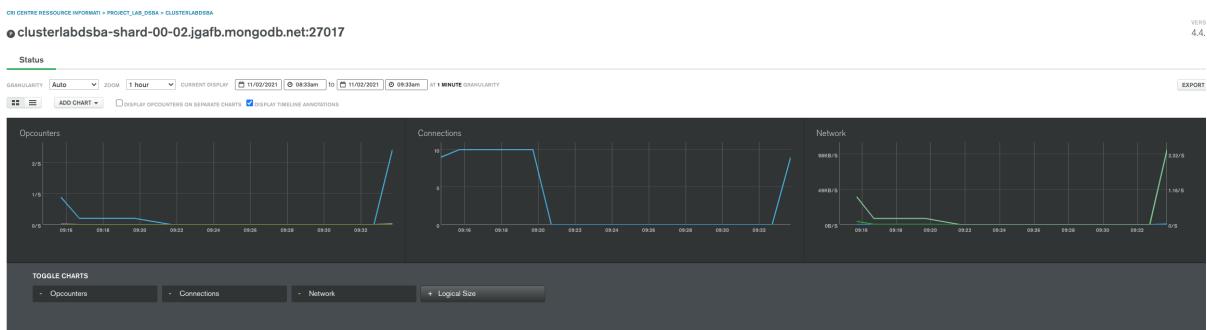
Check the cluster shards :



Investigate for example the primary :



And check how write and/or read queries affect the cluster :



6.11 NOTEBOOK TO CONNECT TO THE CLUSTER

In this part you will interact with your MongoDB database in the cluster. We provided you a first example of connection and access to the **sample_supplies** database but in the following the objective is to repeat the queries you wrote in the first part of the assignment and to check the cluster analysis. The python notebook :

- advanced_spark_notebook_mongodb.ipynb

While you write your queries check how the cluster is interacting.

6.12 FROM THE DOCUMENTATION

You can also check the documentation and find concepts we have seen at lesson. For example how to check that :

- “Atlas does not guarantee that host names remain consistent with respect to node types during topology changes.”