



ST LECTUREs
小 孫 學 堂

Big Data

大數據理論及實務應用


Created by 孫善堂 【小孫學堂】



MongoDB雲端資料庫 串聯pymongo

Created by 孫善堂 【小孫學堂】

創建雲端資料庫

 mongoDB Atlas

All Clusters

Please set your time zone Usage This Month:\$0.00 [details](#) Jason

CONTEXT

TEST1

NON > TEST1

Clusters

Find a cluster...

ATLAS

Clusters

Data Lake BETA

SECURITY

Database Access

Network Access

Advanced

PROJECT

Access Management

Activity Feed

Alerts 0

Integrations


Settings

SERVICES

Charts

Stitch

Triggers



Create a cluster

Choose your cloud provider, region, and specs.

Build a Cluster

Once your cluster is up and running, live migrate an existing MongoDB database into Atlas with our [Live Migration Service](#).

創建雲端資料庫



MONGODB ATLAS

Choose a path. Adjust anytime.

Available as a fully managed service across 60+ regions on AWS, Azure, and Google Cloud

Starter Clusters

For teams learning MongoDB or developing small applications.

- ✓ Highly available auto-healing cluster
- ✓ End-to-end encryption
- ✓ Role-based action control
- ✗ No downtime scaling
- ✗ Network isolation
- ✗ Realtime performance metrics

Starting at

FREE

Create a cluster

Single-Region Clusters

For teams building applications that need advanced development and production-ready environments.

- ✓ Includes all features from Starter Clusters
- ✓ No downtime scaling
- ✓ Network isolation
- ✓ Realtime performance metrics

Starting at

\$0.08/hr*

*estimated cost \$56.94/month

Create a cluster

Multi-Region Clusters

For teams developing world-class applications that require multi-region resiliency or ultra-low latency.

- ✓ Includes all features from Starter and Single-Region Clusters

- ✓ Replicate data across multiple regions

Global Clusters



- ✓ Globally distributed read and write operations
- ✓ Control data residency at the document level

Starting at

\$0.13/hr*

*estimated cost \$98.55/month

Create a cluster

創建雲端資料庫

CLUSTERS > CREATE A STARTER CLUSTER

Create a Starter 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](#).

Cloud Provider & Region

AWS, N. Virginia (us-east-1) ▾



Create a free tier cluster by selecting a region with **FREE TIER AVAILABLE** and choosing the M0 cluster tier below.

★ Recommended region ⓘ

NORTH AMERICA

🇺🇸 N. Virginia (us-east-1) ★

FREE TIER AVAILABLE

🇺🇸 Oregon (us-west-2) ★

FREE TIER AVAILABLE

EUROPE

🇮🇪 Ireland (eu-west-1) ★

FREE TIER AVAILABLE

🇩🇪 Frankfurt (eu-central-1) ★

FREE TIER AVAILABLE

AUSTRALIA

🇦🇺 Sydney (ap-southeast-2) ★

FREE TIER AVAILABLE

ASIA

🇸🇬 Singapore (ap-southeast-1)

★

FREE TIER AVAILABLE

🇮🇳 Mumbai (ap-south-1)

FREE TIER AVAILABLE

Cluster Tier

M0 Sandbox (Shared RAM, 512 MB Storage) ▸

FREE

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

Back

Create Cluster

創建使用者

mongoDB Atlas All Clusters

CONTEXT
TEST1

ATLAS
Clusters
Data Lake BETA
SECURITY
Database Access
Network Access
Advanced

PROJECT
Access Management
Activity Feed
Alerts
Integrations
Settings

SERVICES
Charts
Stitch
Triggers

NON > TEST1

Clusters

Find a cluster...

SANDBOX

Cluster0
Version 4.2.2

CONNECT METRICS COLLECTIONS

CLUSTER TIER
M0 Sandbox (General)

REGION
AWS / N. Virginia (us-east-1)

TYPE
Replica Set - 3 nodes

LINKED STITCH APP
None Linked

Operations R: 0 W: 0
Last 6 Hours

Connections 0
Last 6 Hours

創建使用者

CONTEXT

NON > TEST1

TEST1

Database Access

ATLAS

Clusters

Data Lake BETA

SECURITY

Database Access

Network Access

Advanced

PROJECT

Access Management

Activity Feed

Alerts 0

Integrations

Settings

MongoDB Users

MongoDB Roles

+ ADD NEW USER

User Name

Authentication Method

MongoDB Roles

Actions



Create a database user

Set up database users, permissions, and authentication credentials in order to connect to your clusters.

創建使用者

Add New User

Choose Method

PASSWORD

CERTIFICATE

SCRAM Authentication

SCRAM is MongoDB's default authentication method.

test

e.g. new-user_31

....

SHOW

Autogenerate Secure Password

User Privileges

Atlas admin

Read and write to
any database

Only read any
database

Select Custom
Role

[Add Default Privileges](#)

☐ Save as temporary user

Cancel

Add User

設定可用IP

CONTEXT

NON > TEST1

TEST1 ▾

Database Access

ATLAS

Clusters

Data Lake BETA

SECURITY

Database Access

Network Access

Advanced

PROJECT

Access Management

Activity Feed

Alerts 0

Integrations

Settings

MongoDB Users

MongoDB Roles

+ ADD NEW USER

| User Name ▾ | Authentication Method ▾ | MongoDB Roles | Actions |
|-------------|-------------------------|----------------------------|---|
| 🔍 test | SCRAM | readWriteAnyDatabase@admin | EDIT DELETE |

設定可用IP

mongoDB Atlas

All Clusters

⚠ Please set your time zone

Usage This Month:\$0.00 [details](#)

Jason

CONTEXT

TEST1

NON > TEST1

Network Access

IP Whitelist

Peering

Private Endpoint

+ ADD IP ADDRESS

IP Address

Comment

Status

Actions



Whitelist an IP address

Configure which IP addresses can access your cluster.

[Learn more](#)

設定可用IP

嚴格說來，0.0.0.0/0已經不是一個真正意義上的IP地址了。它表示的是這樣一個集合：所有未知的主機和目的網路。

Add Whitelist Entry

Add a whitelist entry using either CIDR notation or a single IP address. [Learn more.](#)

ADD CURRENT IP ADDRESS

ALLOW ACCESS FROM ANYWHERE

Whitelist Entry:

0.0.0.0/0|

Comment:

Optional comment describing this entry

☐ Save as temporary whitelist

Cancel

Confirm

寫入範例數據(collections)

MongoDB Atlas All Clusters ⚠ Please set your time zone Usage This Month:\$0.00 [details](#) Jason

TEST1 NON > TEST1

Clusters

Find a cluster...

SANDBOX

Cluster0
Version 4.2.2

CONNECT METRICS **COLLECTIONS** ...

CLUSTER TIER
M0 Sandbox (General)

REGION
AWS / N. Virginia (us-east-1)

TYPE
Replica Set - 3 nodes

LINKED STITCH APP
None Linked

Operations R: 0 W: 0 100.0/s
Last 6 Hours

Logical Size 512.0 MB max
Last 30 Days

Connections 0 500 max
Last 6 Hours

Enhance Your Experience
For dedicated throughput, richer metrics and enterprise security options, upgrade your cluster now!

[Upgrade](#)

寫入範例數據(collections)

mongoDB Atlas All Clusters ⚠ Please set your time zone Usage This Month:\$0.00 [details](#) Jason ▾

CONTEXT TEST1 ▾

Overview Real Time Metrics Collections Profiler Performance Advisor Command Line Tools

ATLAS

Clusters

Data Lake BETA

SECURITY

Database Access

Network Access

Advanced

PROJECT

Access Management


Activity Feed

Alerts 0

Integrations

Settings

DATABASES: 0 COLLECTIONS: 0 [REFRESH](#)



Interact with your data

Run queries, view metadata about your collections, manages indexes, and interact with your data with full CRUD functionality.

Load a Sample Dataset Add my own data

[More information](#)

寫入範例數據(collections)

mongoDB Atlas

All Clusters

Please set your time zone

Usage This Month:\$0.00 [details](#)

Jason

CONTEXT

TEST1

NON > TEST1 > CLUSTERS

Cluster0

VERSION 4.2.2

Overview

Real Time

Metrics

Collections

Profiler

Performance Advisor

Command Line Tools

DATABASES: 7

COLLECTIONS: 21

REFRESH

+ Create Database

NAMESPACES

sample_airbnb

listingsAndReviews

sample_analytics

sample_geospatial

sample_mflix

sample_supplies

sample_airbnb.listingsAndReviews

COLLECTION SIZE: 89.99MB

TOTAL DOCUMENTS: 5555

INDEXES TOTAL SIZE: 484KB

Find

Indexes

Aggregation

Full Text Search

INSERT DOCUMENT

FILTER {"beds": 2}

Find

Reset

QUERY RESULTS 1-20 OF MANY

_id: "10009999"

串聯pymongo

mongoDB Atlas

All Clusters

⚠ Please set your time zone

Usage This Month: \$0.00 [details](#)

Jason

CONTEXT

NON > TEST1

TEST1

Clusters

Build a New Cluster

ATLAS

Clusters

Data Lake BETA

SECURITY

Database Access

Network Access

Advanced

PROJECT

Access Management

Activity Feed

Alerts 0

Integrations

Settings

Find a cluster...

Sample dataset successfully loaded. Access it in Data Explorer by clicking the Collections button, or with the Mongo Shell.

[VIEW DATA TUTORIAL](#)

Cluster0

Version 4.2.2

CONNECT

METRICS

COLLECTIONS

...

CLUSTER TIER

M0 Sandbox (General)

REGION

AWS / N. Virginia (us-east-1)

TYPE

Replica Set - 3 nodes

LINKED STITCH APP

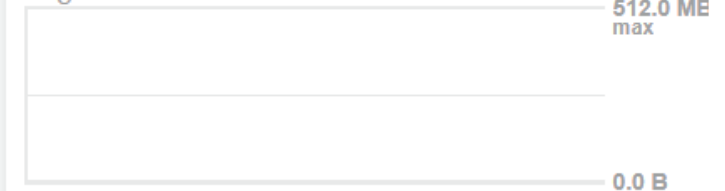
None Linked

Operations R: 0.006 W: 0



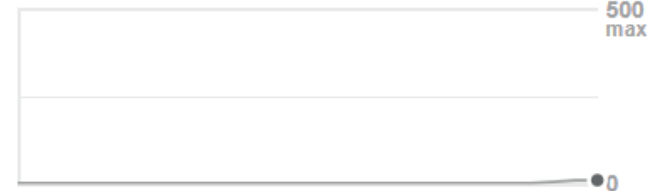
Last 6 Hours

Logical Size



Last 30 Days

Connections 8



Last 6 Hours

Enhance Your Experience

For dedicated throughput, richer metrics and enterprise security options, upgrade your cluster now!

Upgrade

串聯pymongo

Connect to Cluster0

✓ Setup connection security

Choose a connection method

Connect

Choose a connection method [View documentation](#)

See methods to add data and diagnostics in the [Command Line Tools](#) shortcut from within your cluster.



Connect with the Mongo Shell

Mongo Shell with TLS/SSL support is required



Connect Your Application

Get a connection string and view driver connection examples



Connect with MongoDB Compass

Download Compass to explore, visualize, and manipulate your data



串聯pymongo

Connect to Cluster0

✓ Setup connection security

✓ Choose a connection method

Connect

1 Choose your driver version

DRIVER

Python

VERSION

3.6 or later

2 Add your connection string into your application code

Connection String Only

Full Driver Example

```
mongodb+srv://test:<password>@cluster0-2t03u.mongodb.net/test?retryWr:
```

Copy

Replace **<password>** with the password for the **test** user.

When entering your password, make sure that any special characters are [URL encoded](#).

串聯pymongo

```
In [ ]: from pymongo import MongoClient  
client = MongoClient('mongodb+srv://test:<password>@cluster0-2t03u.mongodb.net/t
```

```
In [ ]: from pymongo import MongoClient  
client = MongoClient('mongodb+srv://test:test@cluster0-2t03u.mongodb.net/test?ret
```

Srv錯誤時

ConfigurationError: The "dnspython" module must be installed to use mongodb+srv:// URIs

1. pip3 install dnspython
2. pip3 install pymongo[srv]
3. 記得重開jupyter notebook

timeerror

ServerSelectionTimeoutError

Traceback (most recent call last)

<ipython-input-43-e85f9ec88555> in <module>

As mentioned here: <https://stackoverflow.com/a/54314615/8953378>

26

I added `?ssl=true&ssl_cert_reqs=CERT_NONE` to my connection string, and it fixed the issue.

Timeerror(2)

ServerSelectionTimeoutError

Traceback (most recent call last)

<ipython-input-43-e85f9ec88555> in <module>



Well, I eventually decided to install `certifi` and it worked.

19

```
client = MongoClient(CONNECTION_STRING, tlsCAFile=certifi.where())
```



Wish the docs were a bit clearer on this, but maybe I just didn't look hard enough.

Timeerror(2)

ServerSelectionTimeoutError

Traceback (most recent call last)

<ipython-input-43-e85f9ec88555> in <module>

```
: import certifi
import pymongo

client = pymongo.MongoClient("mongodb+ase?ssl=true&ssl_cert_reqs=CERT_NONE",tlsCAFile=certifi.where())

db=client.sample_airbnb
print(db)
```

串聯pymongo

```
In [43]: import pymongo
```

```
client = pymongo.MongoClient("mongodb+srv://json0523:
```

```
db = client.test
```

```
col=db.student
```

Mongodb架構比較

| SQL | MongoDB |
|------------|----------------|
| database | database |
| table | collection |
| row | document |
| column | field |

使用pymongo

建立/指定資料庫

方法1

```
1 db = client.test
```

方法2

```
1 db = client['test']
```

使用pymongo

建立/指定Collection

方法1

```
1 collection = db.students
```

方法2

```
1 collection = db[students]
```

使用pymongo

新增單筆資料

採`insert_one()`方法

```
1 student = {  
2     'id': '20170101',  
3     'name': 'Jordan',  
4     'age': 20,  
5     'gender': 'male'  
6 }  
7  
8 result = collection.insert_one(student)  
9 print(result)
```

使用pymongo

新增多筆資料

採`insert_many()`方法，傳入一個陣列形式的資料結構

```
1  students = [  
2      {  
3          'id': '20180101',  
4          'name': 'Jean',  
5          'age': 20,  
6          'gender': 'female'  
7      },  
8      {  
9          'id': '20190101',  
10         'name': 'Tom',  
11         'age': 22,  
12         'gender': 'male'  
13     },  
14 ]  
15  
16 result = collection.insert_many(students)  
17 print(result.inserted_ids) # 回傳 ObjectId
```


使用pymongo

可以利用**find_one()**或**find()**方法進行查詢，其中**find_one()**查詢得到的是符合條件的第一個結果。

查詢單一筆資料

如查詢name為Tom的資料

```
1 result = collection.find_one({'name': 'Tom'})
2 print(type(result)) # 回傳結果是dic類型
3 print(result)
```

```
<class 'dict'>
{'_id': ObjectId('5e41253fe2d639ab3976e620'), 'id': '20190101', 'name': 'Tom', 'age': 22, 'gender': 'male'}
```

使用pymongo

查詢多筆資料

```
1 result = list(collection.find({'name': 'Tom'}))  
2 print(result)
```

```
(base) Tsungde-MacBook-Pro:python_practice tsungyuchen$ python mongo_test.py  
[{'_id': ObjectId('5e41253fe2d639ab3976e620'), 'id': '20190101', 'name': 'Tom',  
'age': 22, 'gender': 'male'}, {'_id': ObjectId('5e41262d01b7f8eebd5fc210'), 'id':  
'20190101', 'name': 'Tom', 'age': 22, 'gender': 'male'}]
```

使用pymongo

條件查詢

大於的條件

在dic內使用\$gt作為篩選條件

4.<, <=, >, >= (\$lt, \$lte, \$gt, \$gte)

```
1 results = list(collection.find({'age': {'$gt': 20}}))
2 print(results)
```

```
(base) Tsungde-MacBook-Pro:python_practice tsungyuchen$ python mongo_test.py
[{'_id': ObjectId('5e41253fe2d639ab3976e620'), 'id': '20190101', 'name': 'Tom', 'age': 22, 'gender': 'male'}, {'_id': ObjectId('5e41262d01b7f8eebd5fc210'), 'id': '20190101', 'name': 'Tom', 'age': 22, 'gender': 'male'}]
```

使用pymongo

條件查詢

小於的條件

在dic內使用`$lt`作為篩選條件

```
1 result = list(collection.find({'age': {'$lt': 22}}))
2 print(results)
```

```
(base) Tsungde-MacBook-Pro:python_practice tsungyuchen$ python mongo_test.py
[{'_id': ObjectId('5e412467cd72e1353e0c5421'), 'id': '20170101', 'name': 'Jordan', 'age': 20, 'gender': 'male'}, {'_id': ObjectId('5e41253fe2d639ab3976e61f'), 'id': '20180101', 'name': 'Jean', 'age': 20, 'gender': 'female'}, {'_id': ObjectId('5e41262d01b7f8eebd5fc20f'), 'id': '20180101', 'name': 'Jean', 'age': 20, 'gender': 'female'}]
```


使用pymongo

更新

方法一update()

使用update()方法，如更新name為Tom的age：先指定查詢條件，再查詢該筆資料，修改年齡後呼叫update()方法將原條件和修改後的資料傳入

```
1 condition = {'name': 'Tom'}
2 student = collection.find_one(condition)
3 print(student)
4 student['age'] = 25
5 result = collection.update_one(condition, student)
6 print(result)
```

使用pymongo

更新

也可以使用`$set`對資料進行更新

```
1 result = collection.update(condition, {'$set': student})
```

使用pymongo

更新

方法二: `update_one()`和`update_many()`是官方比較推薦的方法

`update_one()`和`update_many()`是官方比較推薦的方法，用法更加嚴謹

```
1 result = collection.update_one(condition, {'$set': student})
```

嚴謹的方法會回傳UpdateResult object，而`matched_count`和`modified_count`屬性則是指獲得匹配的資料數量和更動的資料數量

```
(base) Tsungde-MacBook-Pro:python_practice tsungyuchen$ python mongo_test.py
<pymongo.results.UpdateResult object at 0x1033cca50>
1 0
```

使用pymongo

刪除

使用`delete_one()`和`delete_many()`方法指定刪除的條件，此時符合條件的資料會被刪除。

```
1 result = collection.delete_many({'name': 'Jean'})
2 print(result)
3 print(result.deleted_count)
4 result = collection.delete_many({'age': {'$gt': 25}})
5 print(result.deleted_count)
```

`deleted_count`為被刪除的數量

```
(base) Tsungde-MacBook-Pro:python_practice tsungyuchen$ python mongo_test.py
<pymongo.results.DeleteResult object at 0x10dd999b0>
1
1
```


使用pymongo

```
In [11]: list(records.find())
```

```
Out[11]: [{ '_id': ObjectId('5ca321051c9d4400003c6883'),  
            'name': 'nikhil',  
            'roll_no': 123,  
            'branch': 'cse'},  
            {'_id': ObjectId('5ca32ed3f9aaa4669a729aeaf'),  
             'name': 'ram',  
             'roll_no': 321,  
             'branch': 'it'},  
            {'_id': ObjectId('5ca32efdfa4aaa4669a729aeb0'),  
             'name': 'alex',  
             'roll_no': 320,  
             'branch': 'it'},  
            {'_id': ObjectId('5ca32efdfa4aaa4669a729aeb1'),  
             'name': 'john',  
             'roll_no': 30,  
             'branch': 'ece'}]
```