



链滴

爬虫数据存储 --MongoDB

作者: [zhui](#)

原文链接: <https://ld246.com/article/1606311146485>

来源网站: 链滴

许可协议: [署名-相同方式共享 4.0 国际 \(CC BY-SA 4.0\)](#)

MongoDB模块

```
pip install pymongo
```

MongoDB链接方式: `mongodb://{mongodb_ip}:{mongodb_port}`

默认端口号: 27017

插入数据

```
import pymongo
```

```
# 创建MongoDB链接
myClient = pymongo.MongoClient("mongodb://127.0.0.1:27017")
# 创建数据库
mydb = myClient['reptiles']
# 创建数据表
myCollection = mydb['reptiles_test']
# 插入单条记录 (数据)
# myCollection.insert_one({"name": "reptiles_test01", "flag": 1})
my_list = [
    {"name": "python", "flag": 100},
    {"name": "shell", "flag": 101},
    {"name": "go", "flag": 102},
    {"name": "html", "flag": 103},
    {"name": "jQuery", "flag": 104}
]
# 插入多条记录
myCollection.insert_many(my_list)
```

注: 新建的数据库表要在插入记录后才可以显示

MongoDB操作:

查看库: **show dbs**

进入库: **use db_name**

查看表: **show collections**

查看数据: **db.table_name.find()**

查看数据

```
# 查看数据
# result = myCollection.find() # find()返回的是一个游标数据, 需要for循环进行查看

# 查看指定字段的数据(只查看name和flag字段)
result = myCollection.find({}, {'_id': 0, 'name': 1, 'flag': 1}) # find({}, {}) 第一个{}代表查询所有,
二个代表限定条件; 0:不显示,1:显示
```

```
for item in result:
    print(item)

# 查看单条数据
# one_result = myCollection.find_one({'name': 'python'})
one_result = myCollection.find_one({'name': 'python'}, {'_id': 0, 'name': 1, 'flag': 1}) # 返回的
数据
print(one_result)

# 高级查询一 [$gt大于; $lt小于]
gt_result = myCollection.find({'flag': {'$gt': 100}}) # 查询flag大于100的数据
for item in gt_result:
    print(item)

# 高级查询二 正则表达式条件查询[$regex]
re_result = myCollection.find({'name': {'$regex': '^p'}})
for item in re_result:
    print(item)
```